PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N — A2

(11) International Publication Number: **WO 97/38085**

(43) International Publication Date: 16 October 1997 (16.10.97)

(21) International Application Number: PCT/US97/05930

(22) International Filing Date: 9 April 1997 (09.04.97)

(30) Priority Data:

60/015,167	9 April 1996 (09.04.96)	US
PCT/US96/09286	5 June 1996 (05.06.96)	WO
(34) Countries for which the regional or international application was filed:		US et al.
60/019,202	6 June 1996 (06.06.96)	US
08/678,280	10 July 1996 (10.07.96)	US

(60) Parent Application or Grant
(63) Related by Continuation
US — 60/015,167 (CIP)
Filed on — 9 April 1996 (09.04.96)

(71) Applicant *(for all designated States except US)*: CALIFORNIA PACIFIC MEDICAL CENTER [US/US]; 2485 Clay Street, San Francisco, CA 94115 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SMITH, Helene [US/US]; 99 Anderson, San Francisco, CA 94110 (US). CHEN, Ling-Chun [US/US]; 510 Lowell Place, Fremont, CA 94536 (US).

(74) Agents: SCHIFF, J., Michael et al.; Morrison & Foerster L.L.P., 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

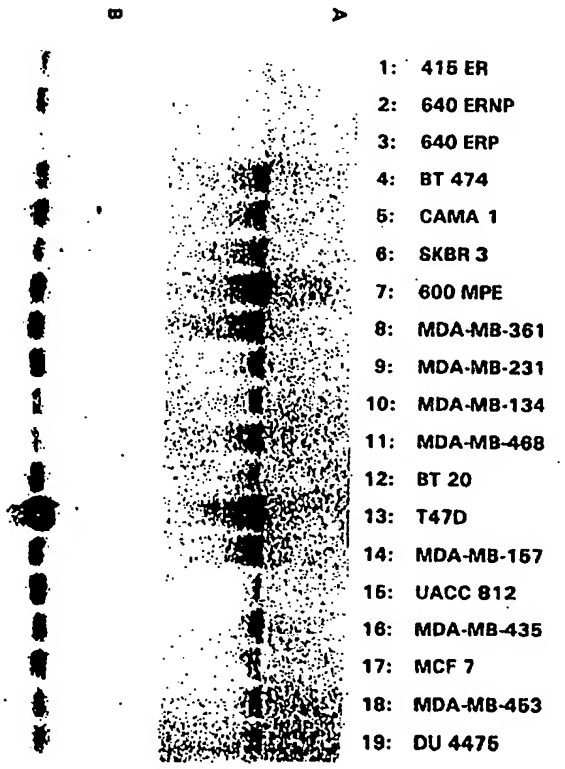
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENES AMPLIFIED IN CANCER CELLS

(57) Abstract

New methods are disclosed for detecting cancer associated genes, and obtaining corresponding cDNA sequences. The methods involve supplying RNA preparations from control cells, and from a plurality of different cancer cells that share a duplicated or deleted gene in the same region of a chromosome. Amplified cDNA copies are displayed, and then selected based on differences in abundance of RNA between preparations. Optional additional screening steps involve surveying panels of cancer cells using the cDNA for RNA overabundance with or without gene duplication. The identified genes can be used in turn to develop materials and techniques for diagnosing and treating the underlying cancer. Four novel genes associated with cancer have been identified. In at least about 60 % of the breast cancer cell lines tested, RNA hybridizing with the cDNAs were substantially more abundant than in normal cells. Most of the cell lines also showed a duplication of the corresponding gene, which probably contributed to the increased level of RNA in the cell. However, for each of the four genes, there were some cell lines which had RNA overabundance without gene duplication. This suggests that the gene product is sufficiently important to the cancer process that cells will use several alternative mechanisms to achieve increased expression.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

GENES AMPLIFIED IN CANCER CELLS

PRIORITY CLAIM

This application claims the priority benefit of the following U.S. Patent applications: 60/015,167, filed April 9, 1996; 60/019,202, filed June 6, 1996; 08/678,280, filed July 10, 1996. For purposes of prosecution in the U.S., the aforementioned applications are hereby incorporated herein by reference in their entirety.

TECHNICAL FIELD

The present invention relates generally to the field of human genetics. More specifically, it relates to the identification of novel genes associated with overabundance of RNA in human cancer such as breast cancer. It pertains especially to those genes and the products thereof which may be important in diagnosis and treatment.

BACKGROUND OF THE INVENTION

Cancer is a heterogeneous disease. It manifests itself in a wide variety of tissue sites, with different degrees of de-differentiation, invasiveness, and aggressiveness. Some forms of cancer are responsive to traditional modes of therapy, but many are not. For most common cancers, there is a pressing need to improve the arsenal of therapies available to provide more precise and more effective treatment in a less invasive way.

As an example, breast cancer has an unsatisfactory morbidity and mortality, despite presently available forms of medical intervention. Traditional clinical initiatives are focused on early diagnosis, followed by surgery and chemotherapy. Such interventions are of limited success, particularly in patients where the tumor has undergone metastasis.

The heterogeneous nature of cancer arises because different cancer cells achieve their growth and pathological properties by different phenotypic alterations. Alteration of gene expression is intimately related to the uncontrolled growth and de-differentiation that are hallmarks of cancer. Certain similar phenotypic alterations in turn may have a different genetic base in different tumors. Yet, the number of genes central to the malignant process must be a finite one. Accordingly, new pharmaceuticals that are tailored to specific genetic alterations in an individual tumor may be more effective.

There are two types of altered gene expression that take place, together or independently, in different cancer cells (reviewed by Bishop). The first type is the decreased expression of recessive genes, known as tumor suppresser genes, that apparently act to prevent malignant growth. The second type is the increased expression of dominant genes, such as oncogenes, that

act to promote malignant growth, or to provide some other phenotype critical for malignancy. Thus, alteration in the expression of either type of gene is a potential diagnostic indicator. Furthermore, a treatment strategy might seek to reinstate the expression of suppresser genes, or reduce the expression of dominant genes. The present invention is directed to identifying genes of either type, particularly those of the second type.

The most frequently studied mechanism for gene overexpression in cancer cells is sometimes referred to as amplification. This is a process whereby the gene is duplicated within the chromosomes of the ancestral cell into multiple copies. The process involves unscheduled replications of the region of the chromosome comprising the gene, followed by recombination of the replicated segments back into the chromosome (Alitalo et al.). As a result, 50 or more copies of the gene may be produced. The duplicated region is sometimes referred to as an "amplicon". The level of expression of the gene (that is, the amount of messenger RNA produced) escalates in the transformed cell in the same proportion as the number of copies of the gene that are made (Alitalo et al.).

Several human oncogenes have been described, some of which are duplicated, for example, in a significant proportion of breast tumors. A prototype is the *erb*B2 gene (also known as HER-2/*neu*), which encodes a 185 kDa membrane growth factor receptor homologous to the epidermal growth factor receptor. *erb*B2 is duplicated in 61 of 283 tumors (22%) tested in a recent survey (Adnane et al.). Other oncogenes duplicated in breast cancer are the *bek* gene, duplicated in 34 out of 286 (12%); the *flg* gene, duplicated in 37 out of 297 (12%), the *myc* gene, duplicated in 43 out of 275 (16%) (Adnane et al.).

Work with other oncogenes, particularly those described for neuroblastoma, suggested that gene duplication of the proto-oncogene was an event involved in the more malignant forms of cancer, and could act as a predictor of clinical outcome (reviewed by Schwab et al. and Alitalo et al.). In breast cancer, duplication of the *erb*B2 gene has been reported as correlating both with reoccurrence of the disease and decreased survival times (Slamon et al.). There is some evidence that *erb*B2 helps identify tumors that are responsive to adjuvant chemotherapy with cyclophosphamide, doxorubicin, and fluorouracil (Muss et al.).

It is clear that only a proportion of the genes that can undergo gene duplication in cancer have been identified. First, chromosome abnormalities, such as double minute (DM) chromosomes and homogeneously stained regions (HSRs), are abundant in cancer cells. HSRs are chromosomal regions that appear in karyotype analysis with intermediate density Giemsa staining throughout their length, rather than with the normal pattern of alternating dark and light bands. They correspond to multiple gene repeats. HSRs are particularly abundant in breast cancers, showing up in 60-65% of tumors surveyed (Dutrillaux et al., Zafrani et al.). When such regions are checked by in situ hybridization with probes for any of 16 known human oncogenes, including *erb*B2 and *myc*, only a proportion of tumors show any hybridization to HSR regions. Furthermore, only a proportion of the HSRs within each karyotype are implicated.

Second, comparative genomic hybridization (CGH) has revealed the presence of copy number increases in tumors, even in chromosomal regions outside of HSRs. CGH is a new method in which whole chromosome spreads are stained simultaneously with DNA fragments from normal cells and from cancer cells, using two different fluorochromes. The images are computer-processed for the fluorescence ratio, revealing chromosomal regions that have undergone amplification or deletion in the cancer cells (Kallioniemi et al. 1992). This method was recently applied to 15 breast cancer cell lines (Kallioniemi et al. 1994). DNA sequence copy number increases were detected in all 23 chromosome pairs.

Cloning the genes that undergo duplication in cancer is a formidable challenge. In one approach, human oncogenes have been identified by hybridizing with probes for other known growth-promoting genes, particularly known oncogenes in other species. For example, the *erb*B2 gene was identified using a probe from a chemically induced rat neuroglioblastoma (Slamon et al.). Genes with novel sequences and functions will evade this type of search. In another approach, genes may be cloned from an area identified as containing a duplicated region by CGH method. Since CGH is able to indicate only the approximate chromosomal region of duplicated genes, an extensive amount of experimentation is required to walk through the entire region and identify the particular gene involved.

Genes may also be overexpressed in cancer without being duplicated. Methods that rely on identification from genetic abnormalities necessarily bypass such genes. Increased expression can come about through a higher level of transcription of the gene; for example, by up-regulation of the promoter or substitution with an alternative promoter. It can also occur if the transcription product is able to persist longer in the cell; for example, by increasing the resistance to cytoplasmic RNase or by reducing the level of such cytoplasmic enzymes. Two examples are the epidermal growth factor receptor, overexpressed in 45% of breast cancer tumors (Klijn et al.), and the IGF-1 receptor, overexpressed in 50-93% of breast cancer tumors (Berns et al.). In almost all cases, the overexpression of each of these receptors is by a mechanism other than gene duplication.

One way of examining overexpression at the messenger RNA level is by subtractive hybridization. It involves producing positive and negative cDNA strands from two RNA preparations, and looking for cDNA which is not completely hybridized by the opposing preparation. This is a laborious procedure which has distinct limitations in cancer research. In particular, since each subtraction involves cDNA from only two cell populations at a time, it is sensitive to individual phenotypic differences due not just to the presence of cancer, but also through natural metabolic variations.

Another way of examining overexpression at the messenger RNA level is by differential display (Liang et al. 1992a). In this technique, cDNA is prepared from only a subpopulation of each RNA preparation, and expanded via the polymerase chain reaction using primers of particular specificity. Similar subpopulations are compared across several RNA preparations by gel autoradiography for expression differences. In order to survey the RNA preparations entirely, the assay is repeated with a comprehensive set of PCR primers. The screening strategy more

effectively includes multiple positive and negative control samples (Sunday et al.). The method has recently been applied to breast cancer cell lines, and highlights a number of expression differences (Liang et al. 1992b; Chen et al., McKenzie et al., Watson et al. 1994 & 1996, Kocher et al.). By excising the corresponding region of the separating gel, it is possible to recover and sequence the cDNA.

Despite the advancement provided by differential display, problems remain in terms of applying it in the search for new cancer genes. First, because this is a test for RNA levels, any phenotypic difference between cell lines constitute part of the recovered set, leading to a large proportion of "false positive" identifications . It has been found that cDNA for mitochondrial genes constitute a large proportion of the differentially expressed bands, and it consumes substantial resources to recover the sample and obtain a partial sequence in order to eliminate them. Second, false positive identifications are made for reasons attributed to multiple cDNA species and competition for the PCR primers by RNA species of different abundance (Debouck). Third, differential display highlights high copy number mRNAs and shorter mRNAs (Bertioli et al., Yeatman et al.) , and may therefore miss critical cancer-associated transcripts when used as a survey technique. Fourth, a number of adjustments are made to gene expression levels when a cell undergoes malignant transformation or cultured in vitro. Most of these adjustments are secondary, and not part of the transformation process. Thus, even when a novel sequence is obtained from the differential display, it is far from certain that the corresponding gene is at the root of the disease process.

An early step in developing gene-specific therapeutic approaches is the identification of genes that are more central to malignant transformation or the persistence of the malignant phenotype.

DISCLOSURE OF THE INVENTION

It is an objective of this invention to provide a method for identifying and characterizing genes and gene products which are duplicated or associated with overabundant RNA in cancer cells. The method can be used for any type of cancer, providing a plurality of cell populations or cell lines of the type of cancer are available, in conjunction with a suitable control cell population. The method is highly effective in identifying genes and gene products that are intimately related to malignant transformation or maintenance of the malignant properties of the cancer cells.

An important derivative of applying the method is the selection and retrieval of cDNA and cDNA fragments corresponding to the cancer-associated gene. These fragments can be used *inter alia* to determine the nucleotide sequence of the gene and mRNA, the amino acid sequence of any encoded protein, or to retrieve from a cDNA or genomic library additional polynucleotides related to the gene or its transcripts. Since the genes are typically involved in the malignant

process of the cell, the polynucleotides, polypeptides, and antibodies derived by using this method can in turn be used to design or screen important diagnostic reagents and therapeutic compounds.

Another objective of this invention to provide isolated polynucleotides, polypeptides, and antibodies derived from four novel genes which are associated with several different types of cancer, including breast cancer. The genes are designated CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1. These designations refer to both strands of the cDNA and fragments thereof; and to the respective corresponding messenger RNA, including splice variants, allelic variants, and fragments of any of these forms. These genes show RNA overabundance in a majority of cancer cell lines tested. A majority of the cells showing RNA overabundance also have duplication of the corresponding gene. Another object of this invention is to provide materials and methods based on these polynucleotides, polypeptides, and antibodies for use in the diagnosis and treatment of cancer, particularly breast cancer.

Accordingly, one embodiment of this invention is an isolated polynucleotide comprising a linear sequence contained in a polynucleotide selected from the group consisting of CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1. The linear sequence is contained in a duplicated gene or overabundant RNA in cancerous cells. The RNA may be overabundant due to gene duplication, increased RNA transcription or processing, increased RNA persistence, any combination thereof, or by any other mechanism, in a proportion of breast cancer cells. Preferably, the RNA is overabundant in at least about 20% of a representative panel of breast cancer cell lines, such as the panels listed herein; more preferably, it is overabundant in at least about 40% of the panel; even more preferably, it is overabundant in at least 60% or more of the panel. Preferably, the RNA is overabundant in at least about 5% of spontaneously occurring breast cancer tumors; more preferably, it is overabundant in at least about 10% of such tumors; more preferably, it is overabundant in at least about 20% of such tumors; more preferably, it is overabundant in at least about 30% of such tumors; even more preferably, it is overabundant in at least about 50% of such tumors.

Preferably, a sequence of at least 10 nucleotides is essentially identical between the isolated polynucleotide of the invention and a cDNA from CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1; more preferably, a sequence of at least about 15 nucleotides is essentially identical; more preferably, a sequence of at least about 20 nucleotides is essentially identical; more preferably, a sequence of at least about 30 nucleotides is essentially identical; more preferably, a sequence of at least about 40 nucleotides is essentially identical; even more preferably, a sequence of at least about 70 nucleotides is essentially identical; still more preferably, a sequence of about 100 nucleotides or more is essentially identical. A further embodiment of this invention is an isolated polynucleotide comprising a linear sequence essentially identical to a sequence selected from the group consisting of SEQ. ID NO:15, SEQ. ID NO:18, SEQ. ID NO:21, SEQ. ID NO:23, SEQ. ID NO:26, SEQ. ID NO:29, SEQ. ID NO:31,. SEQ. ID NO:33, and SEQ. ID NO:35. These embodiments include an isolated polynucleotide which is a DNA polynucleotide, an RNA polynucleotide, a polynucleotide probe, or a polynucleotide primer.

This invention also provides an isolated polypeptide comprising a sequence of amino acids essentially identical to the polypeptide encoded by or translated from a polynucleotide selected from the group consisting of CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1. Preferably, a sequence of at least about 5 amino acids is essentially identical between the polypeptide of this invention and that encoded by the polynucleotide; more preferably, a sequence of at least about 10 amino acids is essentially identical; more preferably, a sequence of at least 15 amino acids is essentially identical; even more preferably, a sequence of at least 20 amino acids is essentially identical; still more preferably, a sequence of about 30 amino acids or more is essentially identical. Preferably, the polypeptide comprises a linear sequence of at least 15 amino acids essentially identical to a sequence encoded by said polynucleotide. Another embodiment of this invention is a polypeptide comprising a linear sequence essentially identical to a sequence selected from the group consisting of SEQ. ID NO:17, SEQ. ID NO:20, SEQ. ID NO:25, SEQ. ID NO:28, SEQ. ID NO:30, SEQ. ID NO:32, SEQ. ID NO:34; and SEQ. ID NO:37.

A further embodiment of this invention is an antibody specific for a polypeptide embodied in this invention. This encompasses both monoclonal and isolated polyclonal antibodies.

A further embodiment of this invention is a method of using the polynucleotides of this invention for detecting or measuring gene duplication in cancerous cells, especially but not limited to breast cancer cells, comprising the steps of reacting DNA contained in a clinical sample with a reagent comprising the polynucleotide, said clinical sample having been obtained from an individual suspected of having cancerous cells; and comparing the amount of complexes formed between the reagent and the DNA in the clinical sample with the amount of complexes formed between the reagent and DNA in a control sample.

A further embodiment is a method of using the polynucleotides of this invention for detecting or measuring overabundance of RNA in cancerous cells, especially but not limited to breast cancer cells, comprising the steps of reacting RNA contained in a clinical sample with a reagent comprising the polynucleotide, said clinical sample having been obtained from an individual suspected of having cancerous cells; and comparing the amount of complexes formed between the reagent and the RNA in the clinical sample with the amount of complexes formed between the reagent and RNA in a control sample.

Another embodiment of this invention is a diagnostic kit for detecting or measuring gene duplication or RNA overabundance in cells contained in an individual as manifest in a clinical sample, comprising a reagent and a buffer in suitable packaging, wherein the reagent comprises a polynucleotide of this invention.

Another embodiment of this invention is a method of using a polypeptide of this invention for detecting or measuring specific antibodies in a clinical sample, comprising the steps of reacting antibodies contained in the clinical sample with a reagent comprising the polypeptide, said clinical sample having been obtained from an individual suspected of having cancerous cells, especially but not limited to breast cancer cells; and comparing the amount of complexes formed between the

reagent and the antibodies in the clinical sample with the amount of complexes formed between the reagent and antibodies in a control sample.

Another embodiment of this invention is a method of using an antibody of this invention for detecting or measuring altered protein expression in a clinical sample, comprising the steps of reacting a polypeptide contained in the clinical sample with a reagent comprising the antibody, said clinical sample having been obtained from an individual suspected of having cancerous cells, especially but not limited to breast cancer cells; and comparing the amount of complexes formed between the reagent and the polypeptide in the clinical sample with the amount of complexes formed between the reagent and a polypeptide in a control sample. Further embodiments of this invention are diagnostic kits for detecting or measuring a polypeptide or antibody present in a clinical sample, comprising a reagent and a buffer in suitable packaging, wherein the reagent respectively comprises either an antibody or a polypeptide of this invention.

Yet another embodiment of this invention is a host cell transfected by a polynucleotide of this invention. A further embodiment of this invention is a method for using a polynucleotide for screening a pharmaceutical candidate, comprising the steps of separating progeny of the transfected host cell into a first group and a second group; treating the first group of cells with the pharmaceutical candidate; not treating the second group of cells with the pharmaceutical candidate; and comparing the phenotype of the treated cells with that of the untreated cells.

This invention also embodies a pharmaceutical preparation for use in cancer therapy, comprising a polynucleotide or polypeptide embodied by this invention, said preparation being capable of reducing the pathology of cancerous cells, especially for but not limited to breast cancer cells. Further embodiments of this invention are methods for treating an individual bearing cancerous cells, such as breast cancer cells, comprising administering any of the aforementioned pharmaceutical preparations.

Still another embodiment of this invention is a pharmaceutical preparation or active vaccine comprising a polypeptide embodied by this invention in an immunogenic form and a pharmaceutically compatible excipient. A further embodiment is a method for treatment of cancer, especially but not limited to breast cancer, either prophylactically or after cancerous cells are present in an individual being treated, comprising administration of the aforementioned pharmaceutical preparation.

Another series of embodiments of this invention relate to methods for obtaining cDNA corresponding to a gene associated with cancer, comprising the steps of: a) supplying an RNA preparation from uncultured control cells; b) supplying RNA preparations from at least two different cancer cells; c) displaying cDNA corresponding to the RNA preparations of step a) and step b) such that different cDNA corresponding to different RNA in each preparation are displayed separately; d) selecting cDNA corresponding to RNA that is present in greater abundance in the cancer cells of step b) relative to the control cells of step a); e) supplying a digested DNA preparation from control cells; f) supplying digested DNA preparations from at least two different cancer cells; g) hybridizing the cDNA of step d) with the digested DNA preparations of step e) and

step f); and h) further selecting cDNA from the cDNA of step d) corresponding to genes that are duplicated in the cancer cells of step f) relative to the control cells of step e).

One or more enhancements may optionally be included in the methods of this invention, including the following:

1. Cancer cells are preferably used for step b) that share a duplicated gene in the same region of a chromosome. If desired, the practitioner may test cancer cells beforehand to detect the duplication or deletion of chromosome regions; or cancer cell lines may be used that have already been characterized in this respect.
2. A higher plurality of cancer cells are preferably used to provide DNA for step b), step f), or preferably both step b) and step f). The use of three cancer cells is preferred over two; the use of four cancer cells is more preferred, about five cancer cells is still more preferred, about eight cancer cells is even more preferred. The cDNA of each cancer cell population is displayed or hybridized separately, in accordance with the method.
3. A higher plurality of control cells are preferably used to provide DNA for step a), step e), or preferably both step a) and step e). The use of two control cell populations is preferred; the use of three or more is even more preferred. Both proliferating and non-proliferating populations are preferably used, if available.
4. The control cells are preferably supplied fresh from a tissue source, and are not cultured or transformed into a cell line. This is increasingly important when the control cell populations used in step a) is only one or two in number. Freshly obtained cancer cells may also be used as an alternative to cancer cell lines, although this is less critical.
5. An additional screening step is preferably conducted in which the cDNA corresponding to the putative cancer-associated gene is additionally hybridized with a digested mitochondrial DNA preparation, to eliminate mitochondrial genes. This screening step may be conducted before, between, subsequent to, or simultaneously with the other screening steps of the method.
6. An additional screening step is preferably conducted in which RNA is supplied from a plurality of cancer cells, and one or preferably more control cell populations; the RNA is contacted with cDNA corresponding to the putative cancer-associated gene under conditions that permit formation of a stable duplex, and cDNA is selected corresponding to RNA that is present in greater abundance in a proportion of the cancer cells relative to the control cells. Preferably, the plurality of cancer cells is a panel of at least five, preferably at least ten cells. Preferably at least three, more preferably at least five of the cancer cells show greater abundance of RNA. Preferably at least one and preferably more of the cancer cells shows a greater abundance of RNA compared with control cells, but does not show duplication of the corresponding gene in step h) of the method.

Other embodiments of the invention are methods for obtaining cDNA corresponding to a gene that is deleted or underexpressed in cancer, comprising the steps of: a) supplying an RNA preparation from control cells; b) supplying RNA preparations from at least two different cancer cells that share a deleted gene in the same region of a chromosome; c) displaying cDNA corresponding to the RNA preparations of step a) and step b) such that different cDNA corresponding to different RNA in each preparation are displayed separately; and d) selecting cDNA corresponding to RNA that is present in lower abundance in the cancer cells of step b) relative to the control cells of step a). Such methods typically comprise the following further steps: e) supplying a digested DNA preparation from control cells; f) supplying digested DNA preparations from at least two different cancer cells; g) hybridizing the cDNA of step d) with the digested DNA preparations of step e) and step f); and h) further selecting cDNA from the cDNA of step d) corresponding to a gene that is deleted in the cancer cells of step f) relative to the control cells of step e). Such methods for identifying deleted or underexpressed genes may also comprise enhancements such as those described above.

Additional embodiments of this invention are methods for characterizing cancer genes, comprising obtaining cDNA corresponding to a cancer-associated gene according to a method of this invention, particularly those highlighted above, and then sequencing the cDNA. Alternatively or in addition, the cDNA may be used to rescue additional polynucleotides corresponding to a cancer-associated gene from an mRNA preparation, or a cDNA or genomic DNA library.

Additional embodiments of this invention are methods for screening candidate drugs for cancer treatment, comprising obtaining cDNA corresponding to a gene that is duplicated, overexpressed, deleted, or underexpressed in cancer, and comparing the effect of the candidate drug on a cell genetically altered with the cDNA or fragment thereof with the effect on a cell not genetically altered.

Various embodiments of this invention may be employed in pursuit of any form of cancer for which suitable tissue sources are available. Cancers of particular interest include lung cancer, glioblastoma, pancreatic cancer, colon cancer, prostate cancer, hepatoma, myeloma, and breast cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a half-tone reproduction of an autoradiogram of a differential display experiment, in which radiolabeled cDNA corresponding to a subset of total messenger RNA in different cells are compared. This is used to select cDNA corresponding to particular RNA that are overabundant in breast cancer.

Figure 2 is a half-tone reproduction of an autoradiogram of electrophoresed DNA digests from a panel of breast cancer cell lines probed with a CH8-2a13-1 insert (Panel A) or a loading control (Panel B).

Figure 3 is a half-tone reproduction of an autoradiogram of electrophoresed total RNA from a panel of breast cancer cell lines probed with a CH8-2a13-1 insert (Panel A) or a loading control (Panel B).

Figure 4 is a half-tone reproduction of an autoradiogram of electrophoresed DNA digests from a panel of breast cancer cell lines probed with a CH13-2a12-1 insert.

Figure 5 is a half-tone reproduction of an autoradiogram of electrophoresed total RNA from a panel of breast cancer cell lines probed with a CH13-2a12-1 insert.

Figure 6 is a map of cDNA fragments obtained for the breast cancer associated genes CH1-9a11-2, CH8-2a13-1, CH13-2a12-1 and CH14-2a16-1. Regions of the fragments used to deduce sequence data listed in the application are indicated by shading. Nucleotide positions are numbered from the left-most residue for which double-strand sequence data has been obtained, which is not necessarily the 5' terminus of the corresponding message.

Figure 7 is a listing of primers used for obtaining the cDNA sequence data for CH1-9a11-2.

Figure 8 is a listing of cDNA sequence obtained for CH1-9a11-2.

Figure 9 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH1-9a11-2 shown in Figure 8. The single-letter amino acid code is used. Stop codons are indicated by a dot (•). The upper panel shows the complete amino acid translation; the lower panel shows the predicted gene product protein sequence. A possible transmembrane region is indicated by underlining.

Figure 10 is a listing of primers used for obtaining the cDNA sequence data for CH8-2a13-1.

Figure 11 is a listing of cDNA sequence obtained for CH8-2a13-1.

Figure 12 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH8-2a13-1 shown in Figure 11. The upper panel shows the complete amino acid translation; the lower panel shows the predicted gene product protein sequence.

Figure 13 is a listing of the nucleotide sequence predicted for a full-length CH8-2a13-1 cDNA.

Figure 14 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH8-2a13-1 shown in Figure 13.

Figure 15 is a listing of primers used for obtaining the cDNA sequence data for CH13-2a12-1.

Figure 16 is a listing of cDNA sequence obtained for CH13-2a12-1.

Figure 17 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH13-2a12-1 shown in Figure 16. The upper panel shows the complete amino acid translation; the lower panel shows the predicted gene product protein sequence.

Figure 18 is a listing of primers used for obtaining cDNA sequence data for CH13-2a12-1..

Figure 19 is a listing of the cDNA sequence data obtained by two-directional sequencing for CH14-2a16-1.

Figure 20 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH14-2a16-1 shown in Figure 19. The upper panel shows the complete amino acid translation; the lower panel shows the predicted gene product protein sequence. Residues corresponding to three zinc finger motifs are underlined, indicating that the protein may have DNA or RNA binding activity.

Figure 21 is a listing of additional DNA sequence data towards the 5' end of CH14-2a16-1 obtained by one-directional sequencing of the fragment pCH14-1.3. First two panels show nucleotide and amino acid sequence from the 5' end of the fragment; the second two panels show nucleotide and amino acid sequence from the 3' end of the fragment. Regions of overlap with pCH14-800 are underlined.

Figure 22 is a listing of the nucleotide sequences of initial fragments obtained corresponding to the four breast cancer associated genes, along with their amino acid translations.

Figure 23 is a listing of additional cDNA sequence obtained for CH1-9a11-2, comprising approximately 1934 base pairs 5' from the sequence of Figure 8.

Figure 24 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH1-9a11-2 shown in Figure 23. The single-letter amino acid code is used. Stop codons are indicated by a dot (●).

Figure 25 is a listing of additional cDNA sequence obtained for CH14-2a16-1, comprising approximately 1934 base pairs 5' from the sequence of Figure 19.

Figure 26 is a listing of the amino acid sequence corresponding to the longest open reading frame of the DNA sequence of CH1-9a11-2 shown in Figure 25. The single-letter amino acid code is used. Stop codons are indicated by a dot (●). The upper panel shows the complete amino acid translation; the lower panel shows zthe predicted gene product protein sequence.

BEST MODE FOR CARRYING OUT THE INVENTION

This invention relates to the discovery and characterization of four novel genes associated with breast cancer. The cDNA of these genes, and their sequences as disclosed below, provide the basis of a series of reagents that can be used in diagnosis and therapy.

Using a panel of about 15 cancer cell lines, each of the four genes was found to be duplicated in 40-60% of the cells tested. Surprisingly, each of the four genes was duplicated in at least one cell line where studies using comparative genomic hybridization had not revealed any amplification of the corresponding chromosomal region.

Levels of expression at the mRNA level were tested in a similar panel for two of these four genes. In addition to those cell lines showing gene duplication, 17 to 37% of the lines showed RNA overabundance without gene duplication, indicating that the malignant cells had used some mechanism *other than gene duplication* to promote the abundance of RNA corresponding to these genes. All four of the breast cancer genes have open reading frames, and likely are transcribed at various levels in different cell types. Overabundance of the corresponding RNA in a cancerous cell is likely associated with overexpression of the protein gene product. Such overexpression may be manifest as increased secretion of the protein from the cell into blood or the surrounding environment, an increased density of the protein at the cell surface, or an increased accumulation the protein within the cell, in comparison to the typical level in noncancerous cells of the same tissue type.

Different tumors bear different genotypes and phenotypes, even when derived from the same tissue. Gene therapy in cancer is more likely to be effective if it is aimed at genes that are involved in supporting the malignancy of the cancer. This invention discloses genes that achieve RNA overabundance by several mechanisms, because they are more likely to be directly involved in the pathogenic process, and therefore suitable targets for pharmacological manipulation.

Features of the four novel genes, the respective mRNA, and the cDNA used to find them are provided in Table 1.

TABLE 1: Characteristics of 4 Novel Breast Cancer Genes

Chromosome	Designation	mRNA Observed	Exemplary cDNA Fragments Cloned
1	CH1-9a11-2	5.5kb, 4.5kb	1.1 kb, 2.5 kb
8	CH8-2a13-1	4.2kb	0.6 kb (two), 3.0 kb, 4.0 kb
13	CH13-2a12-1	3.5kb, 3.2kb	1.6 kb, 3.5 kb
14	CH14-2a16-1	3.8kb, 3kb	0.8 kb, 1.3 kb,1.6 kb, 2.5 kb

All four genes sequences are unrelated to other genes known to be overexpressed in breast cancer, including the *erb*B2 gene (Adnane et al.), tissue factor (Chen et al.), mammaglobulin (Watson et al.), and *DD96* (Kocher et al.).

The four mRNA sequences each comprise an open reading frame. The CH1-9a11-2 gene is expressed at the mRNA level at relatively elevated levels in pancreas and testis. The CH8-2a13-1 gene is expressed at relatively elevated levels in adult heart, spleen, thymus, small intestine, colon, and tissues of the reproductive system; and at higher levels in certain tissues of the fetus. The CH13-2a12-1 gene is expressed at relatively elevated leves in heart, skeletal muscle, and testis. The CH14-2a16-1 gene is expressed at relatively elevated levels in testis. The level of expression of all four genes is especially high in a substantial proportion of breast cancer cell lines.

The CH1-9a11-2 gene encodes a protein with a putative transmembrane region, and may be expressed as a surface protein on cancer cells. The CH13-2a12-1 gene is distantly related to a *C. elegans* gene implicated in cell cycle regulation, and may play a role in the regulation of cell proliferation. The protein encoded by CH13-2a12-1 is distantly related to a vasopressin-activated calcium binding receptor, and may have Ca^{++} binding activity. The CH14-2a16-1 comprises at least five domains of a zinc finger binding motif and is distantly related to a yeast RNA binding protein. The CH14-2a16-1 gene product is suspected of having DNA or RNA binding activity, which may relate to a role in cancer pathogenesis.

The four genes described here are exemplars of genes that undergo altered expression in cancer, identifiable using the gene screening methods of the invention. The method involves an analysis for both DNA duplication and altered RNA abundance relating to the same gene. Since abnormal gene regulation is central to the malignant process, the identification method may be brought to bear on any type of cancer.

The screening method is superior to any previously available approach in several respects. Particularly significant is that screening is rapidly focused towards genes that are central to the malignant process, and away from those that have variable levels of expression as part of normal

metabolic processes. Furthermore, because the end-product is a cDNA corresponding to the gene, the process leads rapidly to detailed characterization of the gene, and any effector molecule it may encode. This in turn leads to development of new diagnostic and therapeutic materials and techniques.

Definitions

Terms used in this application include the following:

The term "polynucleotide" refers to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

The term polynucleotide, as used herein, refers interchangeably to double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of the invention described herein that is a polynucleotide encompasses both the double-stranded form, and each of two complementary single-stranded forms known or predicted to make up the double-stranded form.

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polynucleotide. A "partial sequence" is a linear sequence of part of a polynucleotide which is known to comprise additional residues in one or both directions.

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding is sequence-specific, and typically occurs by Watson-Crick base pairing. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR, or the enzymatic cleavage of a polynucleotide by a ribozyme.

Hybridization reactions can be performed under conditions of different "stringency". Relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. Conditions that increase the stringency of a hybridization reaction are widely known and

published in the art: see, for example, "Molecular Cloning: A Laboratory Manual", Second Edition (Sambrook, Fritsch & Maniatis, 1989).

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. Complementarity (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules.

A linear sequence of nucleotides is "identical" to another linear sequence, if the order of nucleotides in each sequence is the same, and occurs without substitution, deletion, or material substitution. It is understood that purine and pyrimidine nitrogenous bases with similar structures can be functionally equivalent in terms of Watson-Crick base-pairing; and the inter-substitution of like nitrogenous bases, particularly uracil and thymine, or the modification of nitrogenous bases, such as by methylation, does not constitute a material substitution. An RNA and a DNA polynucleotide have identical sequences when the sequence for the RNA reflects the order of nitrogenous bases in the polyribonucleotides, the sequence for the DNA reflects the order of nitrogenous bases in the polydeoxyribonucleotides, and the two sequences satisfy the other requirements of this definition. Where one or both of the polynucleotides being compared is double-stranded, the sequences are identical if one strand of the first polynucleotide is identical with one strand of the second polynucleotide.

A linear sequence of nucleotides is "essentially identical" to another linear sequence, if both sequences are capable of hybridizing to form a duplex with the same complementary polynucleotide. Sequences that hybridize under conditions of greater stringency are more preferred. It is understood that hybridization reactions can accommodate insertions, deletions, and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. In general, essentially identical sequences of about 40 nucleotides in length will hybridize at about 300C in 10 x SSC (0.15 M NaCl, 15 mM citrate buffer); preferably, they will hybridize at about 400C in 6 x SSC; more preferably, they will hybridize at about 500C in 6 x SSC; even more preferably, they will hybridize at about 600C in 6 x SSC, or at about 400C in 0.5 x SSC, or at about 300C in 6 x SSC containing 50% formamide; still more preferably, they will hybridize at 400C or higher in 2 x SSC or lower in the presence of 50% or more formamide. It is understood that the rigor of the test is partly a function of the length of the polynucleotide; hence shorter polynucleotides with the same homology should be tested under lower stringency and longer polynucleotides should be tested under higher stringency, adjusting the conditions accordingly. The relationship between hybridization stringency, degree of sequence identity, and polynucleotide length is known in the art and can be calculated by standard formulae; see, e.g., Meinkoth et al. Sequences that correspond or align more closely to the invention disclosed herein are comparably more preferred. Generally, essentially identical sequences are at least about

50% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 60% identical; more preferably, they are at least about 70% identical; more preferably, they are at least about 80% identical; more preferably, the sequences are at least about 90% identical; even more preferably, they are at least 95% identical; still more preferably, the sequences are 100% identical. Percent identity is calculated as the percent of residues in the sequence being compared that are identical to those in the reference sequence, which is usually one of those listed or described in this application, unless stated otherwise. No penalty is imposed for introduction of gaps in the reference or comparison sequence for purposes of alignment, but the resulting fragments must be rationally derived — small gaps may not be introduced to trivially improve the identity score.

In determining whether polynucleotide sequences are essentially identical, a sequence that preserves the functionality of the polynucleotide with which it is being compared is particularly preferred. Functionality may be established by different criteria, such as ability to hybridize with a target polynucleotide, and whether the polynucleotide encodes an identical or essentially identical polypeptides. Thus, nucleotide substitutions which cause a non-conservative substitution in the encoded polypeptide are preferred over nucleotide substitutions that create a stop codon; nucleotide substitutions that cause a conservative substitution in the encoded polypeptide are more preferred, and identical nucleotide sequences are even more preferred. Insertions or deletions in the polynucleotide that result in insertions or deletions in the polypeptide are preferred over those that result in the down-stream coding region being rendered out of phase. The relative importance of hybridization properties and the polypeptide encoded by a polynucleotide depends on the application of the invention.

A "reagent" polynucleotide, polypeptide, or antibody, is a substance provided for a reaction, the substance having some known and desirable parameters for the reaction. A reaction mixture may also contain a "target", such as a polynucleotide, antibody, or polypeptide that the reagent is capable of reacting with. For example, in some types of diagnostic tests, the amount of the target in a sample is determined by adding a reagent, allowing the reagent and target to react, and measuring the amount of reaction product. In the context of clinical management, a "target" may also be a cell, collection of cells, tissue, or organ that is the object of an administered substance, such as a pharmaceutical compound.

"cDNA" or "complementary DNA" is a single- or double-stranded DNA polynucleotide in which one strand is complementary to a messenger RNA. "Full-length cDNA" is cDNA comprised of a strand which is complementary to an entire messenger RNA molecule. A "cDNA fragment" as used herein generally represents a sub-region of the full-length form, but the entire full-length cDNA may also be included. Unless explicitly specified, the term cDNA encompasses both the full-length form and the fragment form.

Different polynucleotides are said to "correspond" to each other if one is ultimately derived from another. For example, messenger RNA corresponds to the gene from which it is transcribed. cDNA corresponds to the RNA from which it has been produced, such as by a reverse transcription

reaction, or by chemical synthesis of a DNA based upon knowledge of the RNA sequence. cDNA also corresponds to the gene that encodes the RNA. Polynucleotides may be said to correspond even when one of the pair is derived from only a portion of the other.

A "probe" when used in the context of polynucleotide manipulation refers to a polynucleotide which is provided as a reagent to detect a target potentially present in a sample of interest by hybridizing with the target. Usually, a probe will comprise a label or a means by which a label can be attached, either before or subsequent to the hybridization reaction. Suitable labels include, but are not limited to radioisotopes, fluorochromes, chemiluminescent compounds, dyes, and enzymes.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target. A "polymerase chain reaction" ("PCR") is a reaction in which replicate copies are made of a target polynucleotide using one or more primers, and a catalyst of polymerization, such as a reverse transcriptase or a DNA polymerase, and particularly a thermally stable polymerase enzyme. Methods for PCR are taught in U.S. Patent Nos. 4,683,195 (Mullis) and 4,683,202 (Mullis et al.). All processes of producing replicate copies of the same polynucleotide, such as PCR or gene cloning, are collectively referred to herein as "replication."

An "operon" is a genetic region comprising a gene encoding a protein and functionally related 5' and 3' flanking regions. Elements within an operon include but are not limited to promoter regions, enhancer regions, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, protein encoding regions, introns and exons, and termination sites for transcription and translation. A "promoter" is a DNA region capable under certain conditions of binding RNA polymerase and initiating transcription of a coding region located downstream (in the 3' direction) from the promoter. "Operably linked" refers to a juxtaposition of genetic elements, wherein the elements are in a relationship permitting them to operate in the expected manner. For instance, a promoter is operably linked to a coding region if the promoter helps initiate transcription of the coding sequence. There may be intervening residues between the promoter and coding region so long as this functional relationship is maintained.

"Gene duplication" is a term used herein to describe the process whereby an increased number of copies of a particular gene or a fragment thereof is present in a particular cell or cell line. "Gene amplification" generally is synonymous with gene duplication.

"Expression" is defined alternately in the scientific literature either as the transcription of a gene into an RNA polynucleotide, or as the transcription and subsequent translation into a polypeptide. As used herein, "expression" or "gene expression" generally refers to the production of the RNA unless specified or required otherwise. Thus, "RNA overexpression" reflects the presence of more RNA (as a proportion of total RNA) from a particular gene in a cell being described, such as a cancerous cell, in relation to that of the cell it is being compared with, such as a non-cancerous cell. The protein product of the gene may or may not be produced in normal or abnormal amounts. "Protein overexpression" similarly reflects the presence of relatively more protein present in or produced by, for example, a cancerous cell.

"Abundance" of RNA refers to the amount of a particular RNA present in a particular cell type. Thus, "RNA overabundance" or "overabundance of RNA" describes RNA that is present in greater proportion of total RNA in the cell type being described, compared with the same RNA as a proportion of the total RNA in a control cell. A number of mechanisms may contribute to RNA overabundance in a particular cell type: for example, gene duplication, increased level of transcription of the gene, increased persistence of the RNA within the cell after it is produced, or any combination of these. Similarly, "lower abundance" or "underabundance" describes RNA that is present in lower proportion in the cell being described compared with a control cell.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component.

In the context of polypeptides, a "linear sequence" or a "sequence" is an order of amino acids in a polypeptide in an N-terminal to C-terminal direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polypeptide. A "partial sequence" is a linear sequence of part of a polypeptide which is known to comprise additional residues in one or both directions.

A linear sequence of amino acids is "essentially identical" to another sequence if the two sequences have a substantial degree of sequence identity. It is understood that the functional proteins can accommodate insertions, deletions, and substitutions in the amino acid sequence. Thus, linear sequences of amino acids can be essentially identical even if some of the residues do not precisely correspond or align. Sequences that correspond or align more closely to the invention disclosed herein are more preferred. It is also understood that some amino acid substitutions are more easily tolerated. For example, substitution of an amino acid with hydrophobic side chains, aromatic side chains, polar side chains, side chains with a positive or negative charge, or side chains comprising two or fewer carbon atoms, by another amino acid with a side chain of like properties can occur without disturbing the essential identity of the two sequences. Methods for determining homologous regions and scoring the degree of homology are well known in the art; see for example Altschul et al. and Henikoff et al. Well-tolerated sequence differences are referred to as "conservative substitutions". Thus, sequences with conservative substitutions are preferred over those with other substitutions in the same positions; sequences with identical residues at the same positions are still more preferred. In general, amino acid sequences that are essentially identical are at least about 15% identical, and comprise at least about another 15% which are either identical or are conservative substitutions, after alignment of homologous regions. More preferably, essentially identical sequences comprise at least about 50% identical residues or conservative substitutions; more preferably, they comprise at least about 70% identical residues or conservative substitutions; more preferably, they comprise at least about 80% identical residues or conservative substitutions; more

preferably, they comprise at least about 90% identical residues or conservative substitutions; more preferably, they comprise at least about 95% identical residues or conservative substitutions; even more preferably, they contain 100% identical residues.

In determining whether polypeptide sequences are essentially identical, a sequence that preserves the functionality of the polypeptide with which it is being compared is particularly preferred. Functionality may be established by different parameters, such as enzymatic activity, the binding rate or affinity in a receptor-ligand interaction, the binding affinity with an antibody, and X-ray crystallographic structure.

An "antibody" (interchangeably used in plural form) is an immunoglobulin molecule capable of specific binding to a target, such as a polypeptide, through at least one antigen recognition site, located in the variable region of the immunoglobulin molecule. As used herein, the term encompasses not only intact antibodies, but also fragments thereof, mutants thereof, fusion proteins, humanized antibodies, and any other modified configuration of the immunoglobulin molecule that comprises an antigen recognition site of the required specificity.

The term "antigen" refers to the target molecule that is specifically bound by an antibody through its antigen recognition site. The antigen may, but need not be chemically related to the immunogen that stimulated production of the antibody. The antigen may be polyvalent, or it may be a monovalent hapten. Examples of kinds of antigens that can be recognized by antibodies include polypeptides, polynucleotides, other antibody molecules, oligosaccharides, complex lipids, drugs, and chemicals. An "immunogen" is an antigen capable of stimulating production of an antibody when injected into a suitable host, usually a mammal. Compounds may be rendered immunogenic by many techniques known in the art, including crosslinking or conjugating with a carrier to increase valency, mixing with a mitogen to increase the immune response, and combining with an adjuvant to enhance presentation.

An "active vaccine" is a pharmaceutical preparation for human or animal use, which is used with the intention of eliciting a specific immune response. The immune response may be either humoral or cellular, systemic or secretory. The immune response may be desired for experimental purposes, for the treatment of a particular condition, for the elimination of a particular substance, or for prophylaxis against a particular condition or substance.

An "isolated" polynucleotide, polypeptide, protein, antibody, or other substance refers to a preparation of the substance devoid of at least some of the other components that may also be present where the substance or a similar substance naturally occurs or is initially obtained from. Thus, for example, an isolated substance may be prepared by using a purification technique to enrich it from a source mixture. Enrichment can be measured on an absolute basis, such as weight per volume of solution, or it can be measured in relation to a second, potentially interfering substance present in the source mixture. Increasing enrichments of the embodiments of this invention are increasingly more preferred. Thus, for example, a 2-fold enrichment is preferred, 10-fold enrichment is more preferred, 100-fold enrichment is more preferred, 1000-fold enrichment is even more

preferred. A substance can also be provided in an isolated state by a process of artificial assembly, such as by chemical synthesis or recombinant expression.

A polynucleotide used in a reaction, such as a probe used in a hybridization reaction, a primer used in a PCR, or a polynucleotide present in a pharmaceutical preparation, is referred to as "specific" or "selective" if it hybridizes or reacts with the intended target more frequently, more rapidly, or with greater duration than it does with alternative substances. Similarly, an antibody is referred to as "specific" or "selective" if it binds via at least one antigen recognition site to the intended target more frequently, more rapidly, or with greater duration than it does to alternative substances. A polynucleotide or antibody is said to "selectively inhibit" or "selectively interfere with" a reaction if it inhibits or interferes with the reaction between particular substrates to a greater degree or for a greater duration than it does with the reaction between alternative substrates. An antibody is capable of "specifically delivering" a substance if it conveys or retains that substance near a particular cell type more frequently or for a greater duration compared with other cell types.

The "effector component" of a pharmaceutical preparation is a component which modifies target cells by altering their function in a desirable way when administered to a subject bearing the cells. Some advanced pharmaceutical preparations also have a "targeting component", such as an antibody, which helps deliver the effector component more efficaciously to the target site. Depending on the desired action, the effector component may have any one of a number of modes of action. For example, it may restore or enhance a normal function of a cell, it may eliminate or suppress an abnormal function of a cell, or it may alter a cell's phenotype. Alternatively, it may kill or render dormant a cell with pathological features, such as a cancer cell. Examples of effector components are provided in a later section.

A "pharmaceutical candidate" or "drug candidate" is a compound believed to have therapeutic potential, that is to be tested for efficacy. The "screening" of a pharmaceutical candidate refers to conducting an assay that is capable of evaluating the efficacy and/or specificity of the candidate. In this context, "efficacy" refers to the ability of the candidate to effect the cell or organism it is administered to in a beneficial way: for example, the limitation of the pathology of cancerous cells.

A "cell line" or "cell culture" denotes higher eukaryotic cells grown or maintained in vitro. It is understood that the descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell. Cells described as "uncultured" are obtained directly from a living organism, and have been maintained for a limited amount of time away from the organism: not long enough or under conditions for the cells to undergo substantial replication.

"Genetic alteration" refers to a process wherein a genetic element is introduced into a cell other than by mitosis or meiosis. The element may be heterologous to the cell, or it may be an additional copy or improved version of an element already present in the cell. Genetic alteration may be effected, for example, by transfecting a cell with a recombinant plasmid or other polynucleotide through any process known in the art, such as electroporation, calcium phosphate precipitation, or contacting with a polynucleotide-liposome complex, or by transduction or infection

with a DNA or RNA virus or viral vector. The alteration is preferably but not necessarily inheritable by progeny of the altered cell.

A "host cell" is a cell which has been genetically altered, or is capable of being genetically altered, by administration of an exogenous polynucleotide.

The terms "cancerous cell" or "cancer cell", used either in the singular or plural form, refer to cells that have undergone a malignant transformation that makes them pathological to the host organism. Malignant transformation is a single- or multi-step process, which involves in part an alteration in the genetic makeup of the cell and/or the expression profile. Malignant transformation may occur either spontaneously, or via an event or combination of events such as drug or chemical treatment, radiation, fusion with other cells, viral infection, or activation or inactivation of particular genes. Malignant transformation may occur in vivo or in vitro, and can if necessary be experimentally induced.

A frequent feature of cancer cells is the tendency to grow in a manner that is uncontrollable by the host, but the pathology associated with a particular cancer cell may take another form, as outlined infra. Primary cancer cells (that is, cells obtained from near the site of malignant transformation) can be readily distinguished from non-cancerous cells by well-established techniques, particularly histological examination. The definition of a cancer cell, as used herein, includes not only a primary cancer cell, but any cell derived from a cancer cell ancestor. This includes metastasized cancer cells, and in vitro cultures and cell lines derived from cancer cells.

The "pathology" caused by a cancer cell within a host is anything that compromises the well-being or normal physiology of the host. This may involve (but is not limited to) abnormal or uncontrollable growth of the cell, metastasis, release of cytokines or other secretory products at an inappropriate level, manifestation of a function inappropriate for its physiological milieu, interference with the normal function of neighboring cells, aggravation or suppression of an inflammatory or immunological response, or the harboring of undesirable chemical agents or invasive organisms.

"Treatment" of an individual or a cell is any type of intervention in an attempt to alter the natural course of the individual or cell. For example, treatment of an individual may be undertaken to decrease or limit the pathology caused by a cancer cell harbored in the individual. Treatment includes (but is not limited to) administration of a composition, such as a pharmaceutical composition, and may be performed either prophylactically, or subsequent to the initiation of a pathologic event or contact with an etiologic agent. Effective amounts used in treatment are those which are sufficient to produce the desired effect, and may be given in single or divided doses.

A "control cell" is an alternative source of cells or an alternative cell line used in an experiment for comparison purposes. Where the purpose of the experiment is to establish a base line for gene copy number or expression level, it is generally preferable to use a control cell that is not a cancer cell.

The term "cancer gene" as used herein refers to any gene which is yielding transcription or translation products at a substantially altered level or in a substantially altered form in cancerous cells

compared with non-cancerous cells, and which may play a role in supporting the malignancy of the cell. It may be a normally quiescent gene that becomes activated (such as a dominant proto-oncogene), it may be a gene that becomes expressed at an abnormally high level (such as a growth factor receptor), it may be a gene that becomes mutated to produce a variant phenotype, or it may be a gene that becomes expressed at an abnormally low level (such as a tumor suppresser gene). The present invention is directed towards the discovery of genes in all these categories.

It is understood that a "clinical sample" encompasses a variety of sample types obtained from a subject and useful in an in vitro procedure, such as a diagnostic test. The definition encompasses solid tissue samples obtained as a surgical removal, a pathology specimen, or a biopsy specimen, tissue cultures or cells derived therefrom and the progeny thereof, and sections or smears prepared from any of these sources. Non-limiting examples are samples obtained from breast tissue, lymph nodes, and tumors. The definition also encompasses blood, spinal fluid, and other liquid sample of biologic origin, and may refer to either the cells or cell fragments suspended therein, or to the liquid medium and its solutes.

The term "relative amount" is used where a comparison is made between a test measurement and a control measurement. Thus, the relative amount of a reagent forming a complex in a reaction is the amount reacting with a test specimen, compared with the amount reacting with a control specimen. The control specimen may be run separately in the same assay, or it may be part of the same sample (for example, normal tissue surrounding a malignant area in a tissue section).

A "differential" result is generally obtained from an assay in which a comparison is made between the findings of two different assay samples, such as a cancerous cell line and a control cell line. Thus, for example, "differential expression" is observed when the level of expression of a particular gene is higher in one cell than another. "Differential display" refers to a display of a component, particularly RNA, from different cells to determine if there is a difference in the level of the component amongst different cells. Differential display of RNA is conducted, for example, by selective production and display of cDNA corresponding thereto. A method for performing differential display is provided in a later section.

A polynucleotide derived from or corresponding to CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, or CH14-2a16-1 is any of the following: the respective cDNA fragments, the corresponding messenger RNA, including splice variants and fragments thereof, both strands of the corresponding full-length cDNA and fragments thereof, and the corresponding gene. Isolated allelic variants of any of these forms are included. This invention embodies any polynucleotide corresponding to CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, or CH14-2a16-1 in an isolated form. It also embodies any such polynucleotide that has been cloned or transfected into a cell line.

When used in referring to the gene screening methods of this invention (such as those outlined in the last paragraph), "displaying cDNA" is any technique in which DNA copies of RNA (not restricted to mRNA) is rendered detectable in a quantitative or relatively quantitative fashion, in that DNA copies present in a relatively greater amount in a first sample compared with a second sample generates a relatively stronger or weaker signal compared with that of the second sample

due to the difference in copy number. Separate display of different cDNA in a preparation (particularly but not limited to cDNA of different size) allows comparison of levels of a particular cDNA between different samples. A preferred method of display is the differential display technique, and enhancements thereupon described in this disclosure and elsewhere.

The term "digested" DNA encompasses DNA (particularly chromosomal DNA) that has been fragmented by any suitable chemical or enzymatic means into fragments conveniently separable by standard techniques, particularly gel electrophoresis. Digestion with a restriction endonuclease specific for a particular nucleotide sequence is preferred.

"Hybridizing" in this context refers to contacting a first polynucleotide with a second polynucleotide under conditions that permit the formation of a multi-stranded polynucleotide duplex whenever one strand of the first polynucleotide has a sequence of sufficient complementarity to a sequence on the second polynucleotide. The duplex may be a long-lived one, such as when one DNA molecule is used as a labeled probe to detect another DNA molecule, that may optionally be bound to a nitrocellulose filter or present in a separating gel. The duplex may also be a shorter-lived one, such as when one DNA molecule is used to prime an amplification reaction of the other DNA molecule, and the amplified product is subsequently detected. The practitioner may alter the conditions of the reaction to alter the degree of complementarity required, as long as sequence specificity remains a determining factor in the reaction.

Unless explicitly indicated or otherwise required by the techniques used, the steps of a method of this invention may be performed in any order, or combined where desired and appropriate. In one example, in the method comprising steps a) through h) that is described above, it is entirely appropriate to conduct steps a) to c) of the method either before or after steps e) to g) of the method, as long as the cDNA ultimately selected fulfills the criteria of both steps d) and step h). In another example, screening against different digested DNA preparations, even if outlined separately, may optionally be done at the same time. All permutations of this kind are within the scope of the invention.

General methods

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, "Molecular Cloning: A Laboratory Manual", Second Edition (Sambrook, Fritsch & Maniatis, 1989), "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984), "Animal Cell Culture" (R.I. Freshney, ed., 1987); the series "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology" (D.M. Weir & C.C. Blackwell, Eds.), "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987), "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., 1987); and "Current Protocols in Immunology" (J.E. Coligan et al., eds., 1991). All patents, patent applications,

articles and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

Features of the cancer gene screening method

The cancer gene screening methods of this invention may be brought to bear to discover novel genes associated with cancer. Exemplars of cancer-associated genes identified by this method are described below. The exemplars were identified using breast cancer cell lines and tissue, but the strategy can be applied to any cancer type of interest.

A central feature of the cancer gene screening method of this invention is to look for both DNA duplication and RNA overabundance relating to the same gene. This feature is particularly powerful in the discovery of new and potentially important cancer genes. While amplicons occur frequently in cancer, the presently available techniques indicate only the broad chromosomal region involved in the duplication event, not the specific genes involved. The present invention provides a way of detecting genes that may be present in an amplicon from a functional basis. Because an early part of the method involves detecting RNA, the method avoids genes that may be duplicated in an amplicon but are quiescent (and therefore irrelevant) in the cancer cells. Furthermore, it recruits active genes from a duplicated region of the chromosome too small to be detectable by the techniques used to describe amplicons.

Near the heart of this approach are several concepts. One is that genes encoding products implicated positively in the malignant process achieve elevated gene expression as a part of malignant transformation. In this context, "gene expression" refers to expression at the RNA transcription level. Most typically, the RNA is in turn be translated into a protein with a particular enzymatic, binding, or regulatory activity which increases after malignant transformation. In a less common example, the RNA may encode or participate as a ribozyme, antisense polynucleotide, or other functional nucleic acid molecule during malignancy. In a third example, RNA expression may be incidental but symptomatic of an important event in transformation.

Another concept is that overexpression, if central to malignant transformation, may be achieved in different tumors by different mechanisms, and that at least one such possible mechanism is gene duplication. Accordingly, a substantial proportion of transformed cells will have an amplicon, or duplicated region of a chromosome, that includes within its compass the overexpressed gene. Other transformed cells may achieve RNA overabundance without gene duplication, such as by increasing the rate of transcription of the gene (e.g., by upregulation of the promoter region), by enhancing transcript promotion or transport, or by increasing mRNA survival.

Thus, the method entails screening at the RNA level, several cancer cell lines or tumors, and several normal cell lines or tissue samples at the same time. RNA are selected that show a *consistent* elevation amongst the cancer cells as compared with normal cells. Additional strategies may be employed in combination with the RNA screening to improve the success rate of the method. One such strategy is to use several cancer cell lines that are all known to have duplicated

genes in the same region of a particular chromosome. Thus, the RNA that emerge from the screen are more likely to represent a deliberate overexpression event, and the overexpressed gene is likely to be within the duplicated region. A supplemental strategy is to use freshly prepared tissue samples rather than cell lines as controls for base-line expression. This avoids selection of genes that may alter their expression level just as a result of tissue culturing. Another supplemental strategy is to conduct an additional level of screening, following identification of shared, overexpressed RNA. The selected RNA are used to screen DNA from suitable cancer cells and normal cells, to ensure that at least a proportion of the cells achieved the overexpression by way of gene duplication.

The strategy for detecting such genes comprises a number of innovations over those that have been used in previous work.

The first part of the method is based on a search for particular RNAs that are overabundant in cancer cells. A first innovation of the method is to compare RNA abundance between control cells and *several different cancer cells or cancer cell lines* of the desired type. The cDNA fragments that emerge in a greater amount in several different cancer lines, but not in control cells, are more likely to reflect genes that are important in disease progression, rather than those that have undergone secondary or coincidental activation. It is particularly preferred to use cancer cells that are known to share a common duplicated chromosomal region.

A second innovation of this method is to supply as control, not RNA from a cell line or culture, but from *fresh tissue samples* of non-malignant origin. There are two reasons for this. First, the tissue will provide the spectrum of expression that is typical to the normal cell phenotype, rather than individual differences that may become more prominent in culture. This establishes a more reliable baseline for normal expression levels. More importantly, the tissue will be devoid of the effects that in vitro culturing may have in altering or selecting particular phenotypes. For example, proto-oncogenes or growth factors may become up-regulated in culture. When cultured cells are used as the control for differential display, these up-regulated genes would be missed.

A third innovation of this method is to undertake a subselection for cDNA corresponding to genes that achieve their RNA overabundance in a substantial proportion of cancer cells by gene duplication. To accomplish this, appropriate cDNA corresponding to overabundant RNA identified in the foregoing steps are used to probe digests of cellular DNA from a panel of different cancer cells, and from normal genomic DNA. cDNA that shows evidence of higher copy numbers in a proportion of the panel are selected for further characterization. An additional advantage of this step is that cDNA corresponding to mitochondrial genes can rapidly be screened away by including a mitochondrial DNA digest as an additional sample for testing the probe. This eliminates most of the false-positive cDNA, which otherwise make up a majority of the cDNA identified.

Thus, the identification of genes yielding products that are present at abnormal levels is accomplished by a method comprised of the following steps.

To identify particular RNA that is overabundant in cancer cells, RNA is prepared from both cancerous and control cells by standard techniques. Cancer-associated genes may affect cellular

metabolism by any one of a number of mechanisms. For example, they may encode ribozymes, anti-sense polynucleotides, DNA-binding polynucleotides, altered ribosomal RNA, and the like. The gene screening methods of this invention may employ a comparison of RNA abundance levels at the total RNA level, not strictly limited to mRNA. However, the vast majority of cancer-associated genes are predicted to encode a protein gene whose up-regulation is closely linked to the metabolic process. For example, the four exemplary breast cancer genes described elsewhere in this application all comprise an open reading frame. Accordingly, a focus on mRNA enriches the selectable pool for candidate cancer-associated genes. Focus towards mRNA can be conducted at any step in the method. It is particularly convenient to use a display method that displays cDNA copied only from mRNA. In this case, whole RNA may be prepared and analyzed from cancer and control cell populations without separating out mRNA.

In terms of the cancer cells used as an RNA source, it is particularly advantageous to use a plurality of cancer cells known to contain a duplicated gene or chromosomal segment in the same region of the chromosome. The duplicated segment need not be the same size in all the cells, nor is it necessary that the number of duplications be the same, so long as there is at least some part of the duplicated segment that is shared amongst all the cancer cells used in the screen. Thus, a minimum of two, and preferably at least three cancer cells are used that are sufficiently characterized to identify a shared duplicated region, and can be used as a source of RNA for the screening test. In contrast, the control cell population will not comprise chromosomal duplications.

Assuming the duplication to be related to the malignancy of the cancer cells, RNA transcribed from the duplicated region is expected to be overabundant compared with that of the control cell. Accordingly, a highly effective strategy is to identify overabundant RNA that is present in *all (or at least several)* of the cancer cell preparations, but none of the control preparations. By using cancer cells that share a duplicated chromosomal region, the RNA comparison will be strongly biased in favor of RNA overabundance transcribed from the shared duplicated region. Since the shared region is optimally only a small segment of a single chromosome, expression differences arising from elsewhere in the genome in one cancer cell or another will not be selected. We have found that this is highly effective in eliminating: a) RNA abundance differences resulting from normal metabolic variations between cells; and/or b) RNA abundance differences related to cancer cell malignancy, but occurring secondarily to malignant transformation. This is important, because it considerably minimizes the chief deficiency in the use of RNA comparison methods, particularly differential display, for the screening of potential cancer genes: namely, the onerous number of false-positives that such techniques generate.

Shared duplicated regions in cancer cells may be identified by a relevant analytical technique, or by reference to such analysis already conducted and published. One approach that has been highly effective in mapping approximate sub-chromosomal locations of duplicated segments is comparative genomic hybridization (CGH). This technique involves extracting, amplifying and labeling DNA from the subject cell; hybridizing to reference metaphase chromosomes treated to remove repetitive sequences; and observing the position of the hybridized

DNA on the chromosomes (WO 93/18186; Gray et al.). The greater the signal intensity at a given position, the greater the copy number of the sequences in the subject cell. Thus, regions showing elevated staining correspond to genes duplicated in the cancer cells, while regions showing diminished staining correspond to genes deleted in the cancer cells. Related techniques which a practitioner in the art will be well aware are methods for preparing and using repeat sequence chromosome-specific nucleic acid probes (US 5,427,932; Weier et al.), methods for staining target chromosomal DNA using labeled nucleic acid fragments in conjunction with blocking fragments complementary to repetitive DNA segments (US 5,447,841; Gray et al.), and methods for detecting amplified or deleted chromosomal regions using a mapped library of labeled polynucleotide probes (US 5,472,842; Stokke et al.). If desired, multiple fluorochromes can be used as labeling agents with CGH and related techniques, to provide a three-color visualization of deleted, normal, and duplicated chromosome abnormalities (Lucas et al.).

The choice of a particular chromosomal mapping approach is irrelevant, especially once knowledge of the duplicated region is known. If the location of the chromosome duplication is already established for a cell line to be used in RNA comparison during the course of the present invention, then it is unnecessary to conduct a mapping technique *de novo*. For example, established cancer cell lines exist for which mapping data is already available in the public domain. Provided in the reference section of this application is a list of over 40 articles in which the locations of duplicated regions in particular cancer cells are described. In the context of the present invention, a plurality of cancer cells is chosen for the screening panel based on such data, so that they share a duplicated chromosomal region. The chromosomal location of a suspected duplication may be confirmed by hybridization analysis, if desired, using a probe specific for the location.

The cancer cells used for RNA comparison are also generally (but not necessarily) derived from the same type of cancer or the same tissue. Using cells derived from the same type of cancer increases the probability that the gene ultimately identified will be common in that type of cancer, and suitable as a type-specific diagnostic marker. Using cells derived from different types of cancer is in effect a search for cancer-related genes that are less tissue specific and more related to the malignant process in general. Both types of genes are of interest for both diagnostic and therapeutic purposes. In one illustration highlighted in Example 1, RNA was screened from the three breast cancer cell lines BT474, SKBR3, and MCF7, which have been determined by CGH or Southern analysis to share a duplicated genetic regions in chromosomes 1, 8, 14, 17, and 20. When the RNA from these cells was displayed, a number of RNA were found to be overabundant in the cancer cells, but not controls (Figure 1). Three RNA overabundant in all three cancer cell lines corresponded to cancer-associated genes located on chromosomes 1, 8, and 14 that are listed in Table 1. The chromosome 13 gene (CH13-2a12-1) was overexpressed in 2 of the 3 cell lines; namely BT474 and SKBR3. Southern analysis subsequently established that the chromosome 13 gene was duplicated in the same two cell lines (Example 6, Table 5).

Selection of the source or sources of control cell RNA is also a matter of some refinement. The control RNA can be derived from in vitro cultures of non-malignant cells, or established cell lines derived from a non-malignant source. However, it is preferable for the control RNA to be obtained directly from normal human tissue of the same type as the cancer cells. This is because most normal cells do not proliferate indefinitely; hence adaptation of a cell into a cell line involves a degree of transformation. The transforming event may, in turn, be shared with that of certain cancer cells, at least at the level of RNA abundance. Hence, comparison of the RNA levels in cancer cells with so-called control cell lines may lead the practitioner to miss genes that are related to malignancy. For convenience, control cells may be maintained in culture for a brief period before the experiment, and even stimulated; however, multiple rounds of cell division are to be avoided if possible. Use of both stimulated and unstimulated cells as controls may help provide RNA patterns corresponding to the normal range of abundance within various metabolic events of the cell cycle. In one illustration highlighted in Example 1, RNA was screened using both proliferating and non-proliferating cells. As stated, the screening of breast cancer RNA is preferably conducted using uncultured normal mammary epithelial cells (termed "organoids") as sources of control RNA. These cells may be obtained from surgical samples resected from healthy breast tissue.

The RNA is preserved until use in the comparison experiment in such a way to minimize fragmentation. To facilitate confirmation experiments, it is useful to use RNA of a reproducible character. For this reason, it is convenient to use RNA that has been obtained from stable cancerous cell lines and/or ready tissue sources, although reproducibility can also be provided by preparing enough RNA so that it can be preserved in aliquots.

For displaying relative overabundance of RNA in the cancer cells, compared with the control cells, many standard techniques are suitable. These would include any form of subtractive hybridization or comparative analysis. Preferred are techniques in which more than two RNA sources are compared at the same time, such as various types of arbitrarily primed PCR fingerprinting techniques (Welsh et al., Yoshikawa et al.). Particularly preferred are differential mRNA display methods and variations thereof, in which the samples are run in neighboring lanes in a separating gel. These techniques are focused towards mRNA by using primers that are specific for the poly-A tail characteristic of mRNA (Liang et al., 1992a; U.S. Patent 5,262,311).

Because many thousands of genes are expressed in the cells of higher organisms at any one time, it is preferable to improve the legibility of the display by surveying only a subset of the RNA at a time. Methods for accomplishing this are known in the art. A preferred method is by using selective primers that initiate PCR replication for a subset of the RNA. Thus, the RNA is first reverse transcribed by standard techniques. Short primers are used for the selection, preferably chosen such that alternative primers used in a series of like assays can complete a comprehensive survey of the mRNA.

In a preferred example, primers can be used for the 3' region of the mRNAs which have an oligo-dT sequence, followed by two other nucleotides (TiNM, where $i \approx 11$, $N \in \{A,C,G\}$, and $M \in$

{A,C,G,T}). Thus, 12 possible primers are required to complete the survey. A random or arbitrary primer of minimal length can then be used for replication towards what corresponds in the sequence to the 5' region of the mRNA. The optimal length for the random primer is about 10 nucleotides. The product of the PCR reaction is labeled with a radioisotope, such as ^{35}S. The labeled cDNA is then separated by molecular weight, such as on a polyacrylamide sequencing gel.

If desired, variations on the differential display technique may be employed. For example, one-base oligo-dT primers may be used (Liang et al., 1993 & 1994), although this is generally less preferred because the display pattern is correspondingly more complex. Selection of primers may be optimized mathematically depending on the number of RNA species in a tissue of interest (Bauer et al.). The method may be adapted for non-denaturing gels, and for use with automatic DNA sequencers (Bauer et al.). Alternative radioisotopes (Trentmann et al.) or fluorochromes (Sun et al.) may be used for labeling the differential display. Differential display may optionally be combined with a ribonuclease protection assay (Yeatman et al.). PCR primers may optionally incorporate a restriction site to facilitate cloning (Linskens et al., Ayala et al.). Using *Taq* polymerase from multiple manufacturers can increase the amount of variation under otherwise identical conditions (Haag et al.). Nested PCR primers may be used in differential display to decrease background created by oligo-dT primers (WO 95/33760). Other variants of the differential display technique are known in the art and described *inter alia* in the references cited in this disclosure. The use of such modifications are within the scope of the present invention, but are not required, as evidenced by the examples described below.

Based on the comparison of relative abundance of RNA, particular RNAs are chosen which are present as a higher proportion of the RNA in cancerous cells, compared with control cells. When using the differential display method, the cDNA corresponding to overabundant RNA will produce a band with greater proportional intensity amongst neighboring cDNA bands, compared with the proportional intensity in the control lanes. Desired cDNAs can be recovered most directly by cutting the spot in the gel corresponding to the band, and recovering the DNAs therefrom. Recovered cDNA can be replicated again for further use by any technique or combination of techniques known in the art, including PCR and cloning into a suitable carrier.

An optional but highly beneficial additional screening step, typically performed subsequently to an RNA comparison as described above, is aimed at identifying genes that are duplicated in a substantial proportion of cancers. This is conducted by using cDNA such as selected from differential display to probe digests of chromosomal DNA obtained from two or more cancerous cells, such as cancer cell lines. Chromosomal DNA from non-cancerous cells that essentially reflects the germ line in terms of gene copy number is used for the control. A preferred source of control DNA in experiments for human cancer genes is placental DNA, which is readily obtainable. The DNA samples are cleaved at sequence-specific sites along the chromosome, most usually with a suitable restriction enzyme into fragments of appropriate size. The DNA can be blotted directly onto a suitable medium, or separated on an agarose gel before blotting. The latter method is preferred, because it enables a comparison of the hybridizing chromosomal restriction

fragment to determine whether the probe is binding to the same fragment in all samples. The amount of probe binding to DNA digests from each of the cancer cells is compared with the amount binding to control DNA.

Because the comparison is quantitative, it is preferable to standardize the measurement internally. One method is to administer a second probe to the same blot, probing for a second chromosomal gene unlikely to be duplicated in the cancer cells. This method is preferred, because it standardizes not only for differences in the amount of DNA provided, but also for differences in the amount transferred during blotting. This can be accomplished by using alternative labels for the two probes, or by stripping the first probe with a suitable eluant before administering the second.

To eliminate cDNA for mitochondrial genes, it is preferable to include in a parallel analysis a mitochondrial DNA preparation digested with the same restriction enzyme. Any cDNA probe that hybridizes to the appropriate mitochondrial restriction fragments can be suspected of corresponding to a mitochondrial gene.

In the initial replication of the RNA, the random primer may bind at any location along the RNA sequence. Thus, the copied and replicated segment may be a fragment of the full-length RNA. Longer cDNA corresponding to a greater portion of the sequence can be obtained, if desired, by several techniques known to practitioners of ordinary skill. These include using the cDNA fragment to isolate the corresponding RNA, or to isolate complementary DNA from a cDNA library of the same species. Preferably, the library is derived from the same tissue source, and more preferably from a cancer cell line of the same type. For example, for cDNA corresponding to human breast cancer genes, a preferred library is derived from breast cancer cell line BT474, constructed in lambda GT10.

Sequences of the cDNA can be determined by standard techniques, or by submitting the sample to commercial sequencing services. The chromosomal locations of the genes can be determined by any one of several methods known in the art, such as in situ hybridization using chromosomal smears, or panels of somatic cell hybrids of known chromosomal composition.

The cDNA obtained through the selection process outlined can then be tested against a larger panel of cancer cell lines and/or fresh tumor cells to determine what proportion of the cells have duplicated the gene. This can be accomplished by using the cDNA as a probe for chromosomal DNA digests, as described earlier. As illustrated in the Example section, a preferred method for conducting this determination is Southern analysis.

The cDNA can also be used to determine what proportion of the cells have RNA overabundance. This can be accomplished by standard techniques, such as slot blots or blots of agarose gels, using whole RNA or messenger RNA from each of the cells in the panel. The blots are then probed with the cDNA using standard techniques. It is preferable to provide an internal loading and blotting control for this analysis. A preferred method is to re-probe the same blot for transcripts of a gene likely to be present in about the same level in all cells of the same type, such as the gene for a cytoskeletal protein. Thus, a preferred second probe is the cDNA for beta-actin.

Using a novel cDNA found by this selection procedure, it is anticipated that essentially all cancer cells showing gene duplication will also show RNA overabundance, but that some will show RNA overabundance without gene duplication.

The practitioner will readily appreciate that the strategies for identifying genes that are duplicated and/or associated with RNA overabundance may be reversed appropriately to screen for genes that are deleted and/or associated with RNA underabundance. The principles are essentially the same. Genes that are frequently down-regulated in cancer (such as tumor suppresser genes) may be down-regulated by different mechanisms in different cells, and a gene with this behavior is more likely to be central to malignant transformation or persistence of the malignant state.

To screen for such down-regulated genes according to the present invention, RNA is prepared from a plurality of tumors or cancer cell lines and the abundance is compared with RNA preparation from control cells. Again, it is highly preferable to use cancer cells that share a deleted gene in the same chromosomal region, in order to focus any differences at the RNA level towards particular alterations in cancer cells and away from normal variations or coincidental changes. The CGH technique may be used to identify deletions in previously uncharacterized cancer cells. As before, cancer cells may be chosen on the basis of previous knowledge of deleted regions; there is no need to conduct methods such as CGH on previously characterized lines. cDNA from the RNA of cancer cells is displayed (preferably by differential display) alongside cDNA copied from (preferably uncultured) control cells, and cDNA is selected that appears to be underrepresented in at least two (preferably more) of the cancer cells compared with the control cells. cDNA thus selected may optionally be further screened against digested DNA preparations, to confirm that the RNA underabundance observed in the cancer cell populations is attributable in at least a proportion of the cells to an actual gene deletion.

As before, the cDNA may be used for sequencing or rescuing additional polynucleotides, in this case not from the cancer cells but from cells containing or expressing the gene at normal levels. Pharmaceuticals based on deleted genes or those associated with underexpressed RNA are typically oriented at restoring or upregulating the gene, or a functional equivalent of the encoded gene product.

The identification of four exemplary cancer associated genes

To identify particular RNA that is overabundant in cancer cells, RNA has been compared between breast cancer cells and control cells. The amount of total cellular RNA was compared using a modified differential display method. Primers were used for the 3' region of the mRNAs which have an oligo-dT sequence, followed by two other nucleotides as described in the previous section. Random or arbitrary primers of about 10 nucleotides were used for replication towards what corresponds in the sequence to the 5' region of the mRNA. The labeled amplification product was then separated by molecular weight on a polyacrylamide sequencing gel.

Particular mRNAs were chosen that were present in a higher proportion of the RNA in cancerous cells, compared with control cells, according to the proportional intensity amongst neighboring cDNA bands. The cDNA was recovered directly from the gel and amplified to provide a probe for screening. Candidate polynucleotides were screened by a number of criteria, including both Northern and Southern analysis to determine if the corresponding genes were duplicated or responsible for to RNA overabundance in breast cancer cells. Sequence data of the polynucleotides was obtained and compared with sequences in GenBank. Novel polynucleotides with the desired expression patterns were used to probe for longer cDNA inserts in a λgt10 library constructed from the breast cancer cell line BT474, which were then sequenced.

Further description of the actual experimental events that occurred during identification of the four exemplary genes, and sequence data for CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1 are provided in the Example section.

Preparation of polynucleotides, polypeptides and antibodies

Polynucleotides based on the cDNA of CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, CH14-2a16-1, can be rescued from cloned plasmids and phage provided as part of this invention. They may also be obtained from breast cancer cell libraries or mRNA preparations, or from normal human tissues such as placenta, by judicious use of primers or probes based on the sequence data provided herein. Alternatively, the sequence data provided herein can be used in chemical synthesis to produce a polynucleotide with an identical sequence, or that incorporates occasional variations.

Polypeptides encoded by the corresponding mRNA can be prepared by several different methods, all of which will be known to a practitioner of ordinary skill. For example, the appropriate strand of the full-length cDNA can be operably linked to a suitable promoter, and transfected into a suitable host cell. The host cell is then cultured under conditions that allow transcription and translation to occur, and the polypeptide is subsequently recovered. Another convenient method is to determine the polynucleotide sequence of the cDNA, and predict the polypeptide sequence according to the genetic code. A polypeptide can then be prepared directly, for example, by chemical synthesis, either identical to the predicted sequence, or incorporating occasional variations.

Antibodies against polypeptides of this invention may be prepared by any method known in the art. For stimulating antibody production in an animal, it is often preferable to enhance the immunogenicity of a polypeptide by such techniques as polymerization with glutaraldehyde, or combining with an adjuvant, such as Freund's adjuvant. The immunogen is injected into a suitable experimental animal: preferably a rodent for the preparation of monoclonal antibodies; preferably a larger animal such as a rabbit or sheep for preparation of polyclonal antibodies. It is preferable to provide a second or booster injection after about 4 weeks, and begin harvesting the antibody source no less than about 1 week later.

Sera harvested from the immunized animals provide a source of polyclonal antibodies. Detailed procedures for purifying specific antibody activity from a source material are known within the

art. Unwanted activity cross-reacting with other antigens, if present, can be removed, for example, by running the preparation over adsorbants made of those antigens attached to a solid phase, and collecting the unbound fraction. If desired, the specific antibody activity can be further purified by such techniques as protein A chromatography, ammonium sulfate precipitation, ion exchange chromatography, high-performance liquid chromatography and immunoaffinity chromatography on a column of the immunizing polypeptide coupled to a solid support.

Alternatively, immune cells such as splenocytes can be recovered from the immunized animals and used to prepare a monoclonal antibody-producing cell line. See, for example, Harrow & Lane (1988), U.S. Patent Nos. 4,491,632 (J.R. Wands et al.), U.S. 4,472,500 (C. Milstein et al.), and U.S. 4,444,887 (M.K. Hoffman et al.)

Briefly, an antibody-producing line can be produced inter alia by cell fusion, or by transfecting antibody-producing cells with Epstein Barr Virus, or transforming with oncogenic DNA. The treated cells are cloned and cultured, and clones are selected that produce antibody of the desired specificity. Specificity testing can be performed on culture supernatants by a number of techniques, such as using the immunizing polypeptide as the detecting reagent in a standard immunoassay, or using cells expressing the polypeptide in immunohistochemistry. A supply of monoclonal antibody from the selected clones can be purified from a large volume of tissue culture supernatant, or from the ascites fluid of suitably prepared host animals injected with the clone.

Effective variations of this method include those in which the immunization with the polypeptide is performed on isolated cells. Antibody fragments and other derivatives can be prepared by methods of standard protein chemistry, such as subjecting the antibody to cleavage with a proteolytic enzyme. Genetically engineered variants of the antibody can be produced by obtaining a polynucleotide encoding the antibody, and applying the general methods of molecular biology to introduce mutations and translate the variant.

Use in diagnosis

Novel cDNA sequences corresponding to genes associated with cancer are potentially useful as diagnostic aids. Similarly, polypeptides encoded by such genes, and antibodies specific for these polypeptides, are also potentially useful as diagnostic aids.

More specifically, gene duplication or overabundance of RNA in particular cells can help identify those cells as being cancerous, and thereby play a part in the initial diagnosis. Increased levels of RNA corresponding to CH1-9a11-2, CH8-2a13-12, CH13-2a12-1, and CH14-2a16-1 are present in a substantial proportion of breast cancer cell lines and primary breast tumors. In addition, preliminary Northern analysis using probes for CH8-2a13-12, CH13-2a12-1, and CH14-2a16-1 indicates that these genes may be duplicated or be associated with RNA overabundance in certain cell lines derived from cancers other than breast cancer, including colon cancer, lung cancer, prostrate cancer, glioma, and ovarian cancer.

For patients already diagnosed with cancer, gene duplication or overabundance of RNA can assist with clinical management and prognosis. For example, overabundance of RNA may be a useful predictor of disease survival, metastasis, susceptibility to various regimens of standard chemotherapy, the stage of the cancer, or its aggressiveness. See generally the article by Blast, U.S. Patent No. 4,968,603 (Slamon et al.) and PCT Application WO 94/00601 (Levine et al.). All of these determinations are important in helping the clinician choose between the available treatment options.

A particularly important diagnostic application contemplated in this invention is the identification of patients suitable for gene-specific therapy, as outlined in the following section. For example, treatment directed against a particular gene or gene product is appropriate in cancers where the gene is duplicated or there is RNA overabundance. Given a particular pharmaceutical that is directed at a particular gene, a diagnostic test specific for the same gene is important in selecting patients likely to benefit from the pharmaceutical. Given a selection of such pharmaceuticals specific for different genes, diagnostic tests for each gene are important in selecting which pharmaceutical is likely to benefit a particular patient.

The polynucleotide, polypeptide, and antibodies embodied in this invention provide specific reagents that can be used in standard diagnostic procedures. The actual procedures for conducting diagnostic tests are extensively known in the art, and are routine for a practitioner of ordinary skill. See, for example, U.S. Patent No. 4,968,603 (Slamon et al.), and PCT Applications WO 94/00601 (Levine et al.) and WO 94/17414 (K. Keyomarsi et al.). What follows is a brief non-limiting survey of some of the known procedures that can be applied.

Generally, to perform a diagnostic method of this invention, one of the compositions of this invention is provided as a reagent to detect a target in a clinical sample with which it reacts. Thus, the polynucleotide of this invention can be used as a reagent to detect a DNA or RNA target, such as might be present in a cell with duplication or RNA overabundance of the corresponding gene. The polypeptide can be used as a reagent to detect a target for which it has a specific binding site, such as an antibody molecule or (if the polypeptide is a receptor) the corresponding ligand. The antibody can be used as a reagent to detect a target it specifically recognizes, such as the polypeptide used as an immunogen to raise it.

The target is supplied by obtaining a suitable tissue sample from an individual for whom the diagnostic parameter is to be measured. Relevant test samples are those obtained from individuals suspected of containing cancerous cells, particularly breast cancer cells. Many types of samples are suitable for this purpose, including those that are obtained near the suspected tumor site by biopsy or surgical dissection, in vitro cultures of cells derived therefrom, blood, and blood components. If desired, the target may be partially purified from the sample or amplified before the assay is conducted. The reaction is performed by contacting the reagent with the sample under conditions that will allow a complex to form between the reagent and the target. The reaction may be performed in solution, or on a solid tissue sample, for example, using histology sections. The formation of the complex is detected by a number of techniques known in the art. For example, the reagent may be supplied with a label and unreacted reagent may be removed from the complex; the amount of

remaining label thereby indicating the amount of complex formed. Further details and alternatives for complex detection are provided in the descriptions that follow.

To determine whether the amount of complex formed is representative of cancerous or non-cancerous cells, the assay result is compared with a similar assay conducted on a control sample. It is generally preferable to use a control sample which is from a non-cancerous source, and otherwise similar in composition to the clinical sample being tested. However, any control sample may be suitable provided the relative amount of target in the control is known or can be used for comparative purposes. Where the assay is being conducted on tissue sections, suitable control cells with normal histopathology may surround the cancerous cells being tested. It is often preferable to conduct the assay on the test sample and the control sample simultaneously. However, if the amount of complex formed is quantifiable and sufficiently consistent, it is acceptable to assay the test sample and control sample on different days or in different laboratories.

A polynucleotide embodied in this invention can be used as a reagent for determining gene duplication or RNA overabundance that may be present in a clinical sample. The binding of the reagent polynucleotide to a target in a clinical sample generally relies in part on a hybridization reaction between a region of the polynucleotide reagent, and the DNA or RNA in a sample being tested.

If desired, the nucleic acid may be extracted from the sample, and may also be partially purified. To measure gene duplication, the preparation is preferably enriched for chromosomal DNA; to measure RNA overabundance, the preparation is preferably enriched for RNA. The target polynucleotide can be optionally subjected to any combination of additional treatments, including digestion with restriction endonucleases, size separation, for example by electrophoresis in agarose or polyacrylamide, and affixed to a reaction matrix, such as a blotting material.

Hybridization is allowed to occur by mixing the reagent polynucleotide with a sample suspected of containing a target polynucleotide under appropriate reaction conditions. This may be followed by washing or separation to remove unreacted reagent. Generally, both the target polynucleotide and the reagent must be at least partly equilibrated into the single-stranded form in order for complementary sequences to hybridize efficiently. Thus, it may be useful (particularly in tests for DNA) to prepare the sample by standard denaturation techniques known in the art.

The minimum complementarity between the reagent sequence and the target sequence for a complex to form depends on the conditions under which the complex-forming reaction is allowed to occur. Such conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and washing procedure. Higher stringency conditions are those under which higher minimum complementarity is required for stable hybridization to occur. It is generally preferable in diagnostic applications to increase the specificity of the reaction, minimizing cross-reactivity of the reagent polynucleotide alternative undesired hybridization sites in the sample. Thus, it is preferable to conduct the reaction under conditions of high stringency: for example, in the presence of high temperature, low salt, formamide, a combination of these, or followed by a low-salt wash.

In order to detect the complexes formed between the reagent and the target, the reagent is generally provided with a label. Some of the labels often used in this type of assay include radioisotopes such as ^{32}P and ^{33}P, chemiluminescent or fluorescent reagents such as fluorescein, and enzymes such as alkaline phosphatase that are capable of producing a colored solute or precipitant. The label may be intrinsic to the reagent, it may be attached by direct chemical linkage, or it may be connected through a series of intermediate reactive molecules, such as a biotin-avidin complex, or a series of inter-reactive polynucleotides. The label may be added to the reagent before hybridization with the target polynucleotide, or afterwards.

To improve the sensitivity of the assay, it is often desirable to increase the signal ensuing from hybridization. This can be accomplished by replicating either the target polynucleotide or the reagent polynucleotide, such as by a polymerase chain reaction. Alternatively, a combination of serially hybridizing polynucleotides or branched polynucleotides can be used in such a way that multiple label components become incorporated into each complex. See U.S. Patent No. 5,124,246 (Urdea et al.).

An antibody embodied in this invention can also be used as a reagent in cancer diagnosis, or for determining gene duplication or RNA overabundance that may be present in a clinical sample. This relies on the fact that overabundance of RNA in affected cells is often associated with increased production of the corresponding polypeptide. Several of the genes up-regulated in cancer cells encode for cell surface receptors Ă for example, *erb*B-2, c-*myc* and epidermal growth factor. Alternatively, the RNA may encode a protein kept inside the cell, or it may encode a protein secreted by the cell into the surrounding milieu.

Any such protein product can be detected in solid tissue samples and cultured cells by immunohistological techniques that will be obvious to a practitioner of ordinary skill. Generally, the tissue is preserved by a combination of techniques which may include cooling, exchanging into different solvents, fixing with agents such as paraformaldehyde, or embedding in a commercially available medium such as paraffin or OCT. A section of the sample is suitably prepared and overlaid with a primary antibody specific for the protein.

The primary antibody may be provided directly with a suitable label. More frequently, the primary antibody is detected using one of a number of developing reagents which are easily produced or available commercially. Typically, these developing reagents are anti-immunoglobulin or protein A, and they typically bear labels which include, but are not limited to: fluorescent markers such as fluorescein, enzymes such as peroxidase that are capable of precipitating a suitable chemical compound, electron dense markers such as colloidal gold, or radioisotopes such as ^{125}I. The section is then visualized using an appropriate microscopic technique, and the level of labeling is compared between the suspected cancer cell and a control cell, such as cells surrounding the tumor area or those taken from an alternative site.

The amount of protein corresponding to the cancer-associated gene may be detected in a standard quantitative immunoassay. If the protein is secreted or shed from the cell in any appreciable amount, it may be detectable in plasma or serum samples. Alternatively, the target protein may be

solubilized or extracted from a solid tissue sample. Before quantitating, the protein may optionally be affixed to a solid phase, such as by a blot technique or using a capture antibody.

A number of immunoassay methods are established in the art for performing the quantitation. For example, the protein may be mixed with a pre-determined non-limiting amount of the reagent antibody specific for the protein. The reagent antibody may contain a directly attached label, such as an enzyme or a radioisotope, or a second labeled reagent may be added, such as anti-immunoglobulin or protein A. For a solid-phase assay, unreacted reagents are removed by washing. For a liquid-phase assay, unreacted reagents are removed by some other separation technique, such as filtration or chromatography. The amount of label captured in the complex is positively related to the amount of target protein present in the test sample. A variation of this technique is a competitive assay, in which the target protein competes with a labeled analog for binding sites on the specific antibody. In this case, the amount of label captured is negatively related to the amount of target protein present in a test sample. Results obtained using any such assay on a sample from a suspected cancer-bearing source are compared with those from a non-cancerous source.

A polypeptide embodied in this invention can also be used as a reagent in cancer diagnosis, or for determining gene duplication or RNA overabundance that may be present in a clinical sample. Overabundance of RNA in affected cells may result in the corresponding polypeptide being produced by the cells in an abnormal amount. On occasion, overabundance of RNA may occur concurrently with expression of the polypeptide in an unusual form. This in turn may result in stimulation of the immune response of the host to produce its own antibody molecules that are specific for the polypeptide. Thus, a number of human hybridomas have been raised from cancer patients that produce antibodies against their own tumor antigens.

To use the polypeptide in the detection of such antibodies in a subject suspected of having cancer, an immunoassay is conducted. Suitable methods are generally the same as the immunoassays outlined in the preceding paragraphs, except that the polypeptide is provided as a reagent, and the antibody is the target in the clinical sample which is to be quantified. For example, human IgG antibody molecules present in a serum sample may be captured with solid-phase protein A, and then overlaid with the labeled polypeptide reagent. The amount of antibody would then be proportional to the label attached to the solid phase. Alternatively, cells or tissue sections expressing the polypeptide may be overlaid first with the test sample containing the antibody, and then with a detecting reagent such as labeled anti-immunoglobulin. The amount of antibody would then be proportional to the label attached to the cells. The amount of antibody detected in the sample from a suspected cancerous source would be compared with the amount detected in a control sample.

These diagnostic procedures may be performed by diagnostic laboratories, experimental laboratories, practitioners, or private individuals. This invention provides diagnostic kits which can be used in these settings. The presence of cancer cells in the individual may be manifest in a clinical sample obtained from that individual as an alteration in the DNA, RNA, protein, or antibodies contained in the sample. An alteration in one of these components resulting from the presence of

cancer may take the form of an increase or decrease of the level of the component, or an alteration in the form of the component, compared with that in a sample from a healthy individual. The clinical sample is optionally pre-treated for enrichment of the target being tested for. The user then applies a reagent contained in the kit in order to detect the changed level or alteration in the diagnostic component.

Each kit necessarily comprises the reagent which renders the procedure specific: a reagent polynucleotide, used for detecting target DNA or RNA; a reagent antibody, used for detecting target protein; or a reagent polypeptide, used for detecting target antibody that may be present in a sample to be analyzed. The reagent is supplied in a solid form or liquid buffer that is suitable for inventory storage, and later for exchange or addition into the reaction medium when the test is performed. Suitable packaging is provided. The kit may optionally provide additional components that are useful in the procedure. These optional components include buffers, capture reagents, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information.

Use in pharmaceutical development

Embodied in this invention are modes of treating subjects bearing cancer cells that have overabundance of the particular RNA described. The strategy used to obtain the cDNAs provided in this invention was deliberately focused on genes that achieve RNA overabundance by gene duplication in some cells, and by alternative mechanisms in other cells. These alternative mechanisms may include, for example, translocation or enhancement of transcription enhancing elements near the coding region of the gene, deletion of repressor binding sites, or altered production of gene regulators. Such mechanisms would result in more RNA being transcribed from the same gene. Alternatively, the same amount of RNA may be transcribed, but may persist longer in the cell, resulting in greater abundance. This could occur, for example, by reduction in the level of ribozymes or protein enzymes that degrade RNA, or in the modification of the RNA to render it more resistant to such enzymes or spontaneous degradation.

Thus, different cells make use of at least two different mechanisms to achieve a single result Ä the overabundance of a particular RNA. This suggests that RNA overabundance of these genes is central to the cancer process in the affected cells. Interfering with the specific gene or gene product would consequently modify the cancer process. It is an objective of this invention to provide pharmaceutical compositions that enable therapy of this kind.

One way this invention achieves this objective is through screening candidate drugs. The general screening strategy is to apply the candidate to a manifestation of a gene associated with cancer, and then determine whether the effect is beneficial and specific. For example, a composition that interferes with a polynucleotide or polypeptide corresponding any of the novel cancer-associated genes described herein has the potential to block the associated pathology when administered to a tumor of the appropriate phenotype. It is not necessary that the mechanism of interference be known;

only that the interference be preferential for cancerous cells (or cells near the cancer site) but not other cells.

A preferred method of screening is to provide cells in which a polynucleotide related to a cancer gene has been transfected. See, for example, PCT application WO 93/08701. A practitioner of ordinary skill will be well acquainted with techniques for transfecting eukaryotic cells, including the preparation of a suitable vector, such as a viral vector; conveying the vector into the cell, such as by electroporation; and selecting cells that have been transformed, such as by using a reporter or drug sensitivity element.

A cell line is chosen which has a phenotype desirable in testing, and which can be maintained well in culture. The cell line is transfected with a polynucleotide corresponding to one of the cancer-associated genes identified herein. Transfection is performed such that the polynucleotide is operably linked to a genetic controlling element that permits the correct strand of the polynucleotide to be transcribed within the cell. Successful transfection can be determined by the increased abundance of the RNA compared with an untransfected cell. It is not necessary that the cell previously be devoid of the RNA, only that the transfection result in a substantial increase in the level observed. RNA abundance in the cell is measured using the same polynucleotide, according to the hybridization assays outlined earlier.

Drug screening is performed by adding each candidate to a sample of transfected cells, and monitoring the effect. The experiment includes a parallel sample which does not receive the candidate drug. The treated and untreated cells are then compared by any suitable phenotypic criteria, including but not limited to microscopic analysis, viability testing, ability to replicate, histological examination, the level of a particular RNA or polypeptide associated with the cells, the level of enzymatic activity expressed by the cells or cell lysates, and the ability of the cells to interact with other cells or compounds. Differences between treated and untreated cells indicates effects attributable to the candidate. In a preferred method, the effect of the drug on the cell transfected with the polynucleotide is also compared with the effect on a control cell. Suitable control cells include untransfected cells of similar ancestry, cells transfected with an alternative polynucleotide, or cells transfected with the same polynucleotide in an inoperative fashion. Optimally, the drug has a greater effect on operably transfected cells than on control cells.

Desirable effects of a candidate drug include an effect on any phenotype that was conferred by transfection of the cell line with the polynucleotide from the cancer-associated gene, or an effect that could limit a pathological feature of the gene in a cancerous cell. Examples of the first type would be a drug that limits the overabundance of RNA in the transfected cell, limits production of the encoded protein, or limits the functional effect of the protein. The effect of the drug would be apparent when comparing results between treated and untreated cells. An example of the second type would be a drug that makes use of the transfected gene or a gene product to specifically poison the cell. The effect of the drug would be apparent when comparing results between operably transfected cells and control cells.

Use in treatment

This invention also provides gene-specific pharmaceuticals in which each of the polynucleotides, polypeptides, and antibodies embodied herein as a specific active ingredient in pharmaceutical compositions. Such compositions may decrease the pathology of cancer cells on their own, or render the cancer cells more susceptible to treatment by the non-specific agents, such as classical chemotherapy or radiation.

An example of how polynucleotides embodied in this invention can be effectively used in treatment is gene therapy. See, for example, Morgan et al., Culver et al., and U.S. Patent No. 5,399,346 (French et al.). The general principle is to introduce the polynucleotide into a cancer cell in a patient, and allow it to interfere with the expression of the corresponding gene, such as by complexing with the gene itself or with the RNA transcribed from the gene. Entry into the cell is facilitated by suitable techniques known in the art as providing the polynucleotide in the form of a suitable vector, or encapsulation of the polynucleotide in a liposome. The polynucleotide may be provided to the cancer site by an antigen-specific homing mechanism, or by direct injection.

A preferred mode of gene therapy is to provide the polynucleotide in such a way that it will replicate inside the cell, enhancing and prolonging the interference effect. Thus, the polynucleotide is operably linked to a suitable promoter, such as the natural promoter of the corresponding gene, a heterologous promoter that is intrinsically active in cancer cells, or a heterologous promoter that can be induced by a suitable agent. Preferably, the construct is designed so that the polynucleotide sequence operably linked to the promoter is complementary to the sequence of the corresponding gene. Thus, once integrated into the cellular genome, the transcript of the administered polynucleotide will be complementary to the transcript of the gene, and capable of hybridizing with it. This approach is known as anti-sense therapy. See, for example, Culver et al. and Roth.

The use of antibodies embodied in this invention in the treatment of cancer partly relies on the fact that genes that show RNA overabundance in cancer frequently encode cell-surface proteins. Location of these proteins at the cell surface may correspond to an important biological function of the cancer cell, such as their interaction with other cells, the modulation of other cell-surface proteins, or triggering by an incoming cytokine.

These mechanisms suggest a variety of ways in which a specific antibody may be effective in decreasing the pathology of a cancer cell. For example, if the gene encodes for a growth receptor, then an antibody that blocks the ligand binding site or causes endocytosis of the receptor would decrease the ability of the receptor to provide its signal to the cell. It is unnecessary to have knowledge of the mechanism beforehand; the effectiveness of a particular antibody can be predicted empirically by testing with cultured cancer cells expressing the corresponding protein. Monoclonal antibodies may be more effective in this form of cancer therapy if several different clones directed at different determinants of the same cancer-associate gene product are used in combination: see PCT application WO 94/00136 (Kasprzyk et al.). Such antibody treatment may directly decrease the

pathology of the cancer cells, or render them more susceptible to non-specific cytotoxic agents such as platinum (Lippman).

Another example of how antibodies can be used in cancer therapy is in the specific targeting of effector components. The protein product of the cancer-associated gene is expected to appear in high frequency on cancer cells compared to unaffected cells, due to the overabundance of the corresponding RNA. The protein therefore provides a marker for cancer cells that a specific antibody can bind to. An effector component attached to the antibody therefore becomes concentrated near the cancer cells, improving the effect on those cells and decreasing the effect on non-cancer cells. This concentration would generally occur not only near the primary tumor, but also near cancer cells that have metastasized to other tissue sites. Furthermore, if the antibody is able to induce endocytosis, this will enhance entry of the effector into the cell interior.

For the purpose of targeting, an antibody specific for the protein of the cancer-associated gene is conjugated with a suitable effector component, preferably by a covalent or high-affinity bond. Suitable effector components in such compositions include radionuclides such as ^{131}I, toxic chemicals such as vincristine, and toxic peptides such as diphtheria toxin. Other suitable effector components include peptides or polynucleotides capable of altering the phenotype of the cell in a desirable fashion: for example, installing a tumor suppresser gene, or rendering them susceptible to immune attack.

In most applications of antibody molecules in human therapy, it is preferable to use human monoclonals, or antibodies that have been humanized by techniques known in the art. This helps prevent the antibody molecules themselves from becoming a target of the host's immune system.

An example of how polypeptides embodied in this invention can be effectively used in treatment is through vaccination. The growth of cancer cells is naturally limited in part due to immune surveillance. This refers to the recognition of cancer cells by immune recognition units, particularly antibodies and T cells, and the consequent triggering of immune effector functions that limit tumor progression. Stimulation of the immune system using a particular tumor-specific antigen enhances the effect towards the tumor expressing the antigen. Thus, an active vaccine comprising a polypeptide encoded by the cDNA of this invention would be appropriately administered to subjects having overabundance of the corresponding RNA. There may also be a prophylactic role for the vaccine in a population predisposed for developing cancer cells with overabundance of the same RNA.

Ways of increasing the effectiveness of cancer vaccines are known in the art (Beardsley, MacLean et al.). For example, synthetic antigens are conjugated to a carrier like keyhole limpet hemocyanin (KLH), and then combined with an adjuvant such as DETOX™, a mixture of mycobacterial cell walls and lipid A. Any polypeptide encoded by the four novel genes described in this invention can be used in analogous compositions.

Methods for preparing and administering polypeptide vaccines are known in the art. Peptides may be capable of eliciting an immune response on their own, or they may be rendered more immunogenic by chemical manipulation, such as cross-linking or attaching to a protein carrier like KLH. Preferably, the vaccine also comprises an adjuvant, such as alum, muramyl dipeptides,

liposomes, or DETOX™. The vaccine may optionally comprise auxiliary substances such as wetting agents, emulsifying agents, and organic or inorganic salts or acids,. It also comprises a pharmaceutically acceptable excipient which is compatible with the active ingredient and appropriate for the route of administration. The desired dose for peptide vaccines is generally from 10 μg to 1 mg, with a broad effective latitude. The vaccine is preferably administered first as a priming dose, and then again as a boosting dose, usually at least four weeks later. Further boosting doses may be given to enhance the effect. The dose and its timing are usually determined by the person responsible for the treatment.

Sequence data and deposits

The foregoing detailed description provides, inter alia, a detailed explanation of how genes associated with cancer can be identified and their cDNA obtained. Polynucleotide sequences for CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1 are provided.

The sequence data listed in this application was obtained by two-directional sequencing, except where indicated otherwise. The data are believed to be accurate — nevertheless, it is readily appreciated that the techniques of the art as used herein have the potential of introducing occasional and infrequent sequence errors. Clones and inserts obtained via PCR may also comprise occasional errors introduced during amplification. Nucleotide sequences predicted from database compilations, and sequence data obtained by one-directional sequencing may also contain occasional errors in accordance with the limitations of the underlying techniques. In addition, allelic variations to both nucleotide and amino acid sequences may occur naturally or be deliberately induced. Differences of any of these types between the sequences provided herein and the invention as practiced may be present without departing from the spirit of the invention.

Sequence data for CH8-2a13-1 and CH13-2a12-1 cDNA are believed to comprise the entire translated coding sequence, and 5' and 3' untranslated regions corresponding to those found in typical mRNA transcripts. Multiple mRNA transcripts may be found depending on the patterns of transcript processing in various cell types of interest. Sequence data for CH1-9a11-2 and CH14-2a16-1 cDNA comprise a portion of the coding sequence and 3' untranslated regions. Additional sequence is typically present in the corresponding mRNA transcripts, comprising an additional coding region in the N-terminal direction of the protein, and possibly a 5' untranslated region.

Certain embodiments of this invention may be practiced by polynucleotide synthesis according to the data provided herein, by rescuing an appropriate insert corresponding to the gene of interest from one of the deposits listed below, or by isolating a corresponding polynucleotide from a suitable tissue source. Various useful probes and primers for use in polynucleotide isolation are provided herein, or may be designed from the sequence data.

Three deposits have been made on May 31, 1996 with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852 under terms of the Budapest treaty. The deposits are outlined in Table 2:

TABLE 2: ATCC Deposits

BCGF1 Accession No. 98074	Mixture of *E. coli* with recombinant plasmids of cDNA fragments of genes associated with breast cancer. The 8 recombinant plasmids may be separated by plating on Ampicillin plates and selecting single colonies for analysis by PCR using SP6 and T7 primers.		
	Gene	Subclone	Expected size of PCR product
	CH1-9a11-2	pch1-1.1	1.1 kb
		pch1-2.5	2.5 kb
	CH8-2a13-1	pch8-600	600 bp
		pch8-3k	3.0 kb
		pch8-4k	4.0 kb
	CH14-2a16-1	pch14-800	800 kb
		pch14-1.6	1.6 kb
		pch14-1.3	1.3 kb
BCGF 2 Accession No. 97595	Mixture of λgt10 recombinant phages with cDNA inserts of genes associated with breast cancer. The 2 phages may be separated by growing in the *E. coli* host (strain NM514) and plating out for single plaques. These plaques can be distinguished by PCR using λgt10 reverse and forward primers.		
	Gene	Phage	Expected size of PCR product
	CH13-2a12-1	λch13-3.5	3.5 kb
	CH14-2a16-1	λch14-2.5	2.5 kb
λBCBT474 Accession No. 97594	cDNA library derived from breast cancer cell line BT474 in λgt10 vector, supplemented with a cDNA library from breast cancer cell line 600PE in λgt10 vector. The cDNA insert sizes range from about 0.5 to 5 kb. λBCBT474 is a source of additional cDNA inserts corresponding to CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, or CH14-2a16-1 not present in BCGF-1 or BCGF-2.		

Sequence databases contain sequences of polynucleotide and polypeptide fragments with varyous degrees of identity and overlap with certain embodiments of this invention. The following list of accession numbers is provided for the interest of the reader; it is not intended to be comprehensive or a limitation on the invention. The database disclosures do not typically indicate use in cancer diagnosis, drug development, or disease treatment.

The following GenBank accession numbers are listed in relation to CH1-9a11-2: dbEST N32686; N45113; N36176; N22982; AA278830; H88670; AA235936; AA236951; H26301; N28026;

H88063; H88064; D61948; H88718; H26460; AA137920; AA145308; W12952; AA200687; N44164; T27279; dbSTS G22044; G04961.

The following GenBank accession numbers are listed in relation to CH8-2a13-1: dbNR D83780

The following GenBank accession numbers are listed in relation to CH13-2a12-1: dbNR U58090; dbEST AA182441; AA253924; AA179755; AA112715; AA112640; W67977; AA150317; W68080; AA150243; AA100446; W69636; H46574; AA245889; AA100651; H77368; AA192778; T85671; N32682; T86257; T78239; T77874; AA187866; Z33557; R40816; N99802; R19302; AA100650; N55904; AA257151; H77369; T79014.

The following GenBank accession numbers are listed in relation to CH14-2a16-1: dbEST N64802; W56903; N31400; W95674; AA233551; AA233636; N24105; W03447; W25821; AA233666; AA233647; N67843; D55778; T66839; N55370; N75650; AA280736; H97110; Z19643; H91250; AA230765; R93089; T84665; W94857; R92873

The examples presented below are provided as a further guide to a practitioner of ordinary skill in the art, and are not meant to be limiting in any way.

EXAMPLES

Example 1: Selecting cDNA for messenger RNA that is overabundant in breast cancer cells

Total RNA was isolated from each breast cancer cell line or control cell by centrifugation through a gradient of guanidine isothiocyanate/CsCl. The RNA was treated with RNase-free DNase (Promega, Madison, WI). After extraction with phenol-chloroform, the RNA preparations were stored at -70°C. Oligo-dT polynucleotides for priming at the 3' end of messenger RNA with the sequence $T_{11}NM$ (where $N \in \{A,C,G\}$ and $M \in \{A,C,G,T\}$) were synthesized according to standard protocols. Arbitrary decamer polynucleotides (OPA01 to OPA20) for priming towards the 5' end were purchased from Operon Biotechnology, Inc., Alameda, CA.

The RNA was reverse-transcribed using AMV reverse transcriptase (obtained from BRL) and an anchored oligo-dT primer in a volume of 20 µL, according to the manufacturer's directions. The reaction was incubated at 370C for 60 min and stopped by incubating at 950C for 5 min. The cDNA obtained was used immediately or stored frozen at -70°C.

Differential display was conducted according to the following procedure: 1 µL cDNA was replicated in a total volume of 10 µL PCR mixture containing the appropriate $T_{11}NM$ sequence, 0.5 TM of a decamer primer, 200 TM dNTP, 5 TCi [^{35}S]-dATP (Amersham), Taq polymerase buffer with 2.5 mM $MgCl_2$ and 0.3 unit Taq polymerase (Promega). Forty cycles were conducted in the following sequence: 94°C for 30 sec, 40°C for 2 min, 72°C for 30 sec; and then the sample was incubated at

72°C for 5 min. The replicated cDNA was separated on a 6% polyacrylamide sequencing gel. After electrophoresis, the gel was dried and exposed to X-ray film.

The autoradiogram was analyzed for labeled cDNA that was present in larger relative amount in all of the lanes corresponding to breast cancer cells, compared with all of the lanes corresponding to control cells. Figure 1 provides an example of an autoradiogram from such an experiment. Lane 1 is from non-proliferating normal breast cells; lane 2 is from proliferating normal breast cells; lanes 3 to 5 are from breast cancer cell lines BT474, SKBR3, and MCF7. The left and right side shows the pattern obtained from experiments using the same T_{11}NM sequence (T_{11}AC), but two different decamer primers. The arrows indicate the cDNA fragments that were more abundant in all three tumor lines compared with controls.

The assay illustrated in Figure 1 was conducted using different combinations of oligo-dT primers and decamer primers. A number of differentially expressed bands were detected when different primer combinations were used. However, not all differences seen initially were reproducible after re-screening. We therefore routinely repeated each differential display for each primer combination. Only bands showing RNA overabundance in at least 2 experiments were selected for further analysis.

It is preferable to include in the differential display experiment RNA derived from uncultured normal mammary epithelial cells (termed "organoids"). These cells are obtained from surgical samples resected from healthy breast tissue, which are then coaxed apart by blunt dissection techniques and mild enzyme treatment. Using organoids as the negative control, 33 cDNA fragments were isolated from 15 displays.

Example 2: Sub-selecting cDNA that corresponds to genes that are duplicated in breast cancer cells

cDNA fragments that were differentially expressed in the fashion described in Example 1 were excised from the dried gel and extracted by boiling at 950C for 10 min. Eluted cDNA was recovered by ethanol precipitation, and replicated by PCR. The product was cloned into the pCRII vector using the TA cloning system (Invitrogen).

*Eco*RI digested placenta DNA, and *Eco*RI digested DNA from the breast cancer cell lines BT474, SKBR3 and ZR-75-30 were used to prepare Southern blots to screen the cloned cDNA fragments. The cloned cDNA fragments were labeled with [32P]-dCTP, and used individually to probe the blots. A larger relative amount of binding of the probe to the lanes corresponding to the cancer cell DNA indicated that the corresponding gene had been duplicated in the cancer cells. The labeled cDNA probes were also used in Northern blots to verify that the corresponding RNA was overabundant in the appropriate cell lines.

To determine whether the cDNA fragments obtained by this selection procedure corresponded to novel genes, a partial nucleotide sequence was obtained using M13 primers. Each sequence was compared with the known sequences in GenBank. In initial experiments, 5 of

the first 7 genes sequenced were mitochondrial genes. To avoid repeated isolation of mitochondrial genes, subsequent screening experiments were done with additional lanes in the DNA blot analysis for *Eco*RI digested and *Hind*III digested mitochondrial DNA. Any cDNA fragment that hybridized to the appropriate mitochondrial restriction fragments was suspected of corresponding to a mitochondrial gene, and not analyzed further.

From the 33 cDNA fragments detected from differential displays using organoid mRNA, 12 were subcloned. Of these 12, 6 detected suitable gene duplications in the appropriate cell lines. Three cDNA failed to detect duplicated genes, and 3 appeared to correspond to mitochondrial genes. Sequence analysis of the 6 suitable cDNA fragments showed no identity to any known genes.

To obtain longer cDNA corresponding to the cDNA fragments with novel sequences, the fragments were used as probes to screen a cDNA library from breast cancer cell line BT474, constructed in lambda GT10. The longer cDNA obtained from lambda GT10 were sequenced using lambda GT10 primers. The chromosomal locations of the cDNAs were determined using panels of somatic cell hybrids.

Four of the 6 novel cDNA identified so far have been processed in this fashion. The probes used to obtain the 4 new breast cancer genes are shown in Table 3.

TABLE 3: Primers used for Differential Display

cDNA	Oligo-dT primer	Arbitrary primer
CH1-9a11-2	$T_{11}CC$ (SEQ ID NO: 9)	SEQ ID NO:11
CH8-2a13-1	$T_{11}AC$ (SEQ ID NO:10)	SEQ ID NO:12
CH13-2a12-1	$T_{11}AC$ (SEQ ID NO:10)	SEQ ID NO:13
CH14-2a16-1	$T_{11}AC$ (SEQ ID NO:10)	SEQ ID NO:14

Example 3: Using the cDNA to test panels of breast cancer cells

To determine the proportion of breast cancers in which the putative breast cancer genes were duplicated, or showed RNA overabundance without gene duplication, the four cDNA obtained according to the selection procedures described were used to probe a panel of breast cancer cell lines and primary tumors.

Gene duplication was detected either by Southern analysis or slot-blot analysis. For Southern analysis, 10 µg of *Eco*RI digested genomic DNA from different cell lines was

electrophoresed on 0.8% agarose and transferred to a HYBOND™ N+ membrane (Amersham). The filters were hybridized with 32P-labeled cDNA for the putative breast cancer gene. After an autoradiogram was obtained, the probe was stripped and the blot was re-probed using a reference probe to adjust for differences in sample loading. Either chromosome 2 probe D2S5 or chromosome 21 probe D21S6 was used as a reference. Densities of the signals on the autoradiograms were obtained using a densitometer (Molecular Dynamics). The density ratio between the breast cancer gene and the reference gene was calculated for each sample. Two samples of placental DNA digests were run in each Southern analysis as a control.

For slot-blot analysis, 1 μg of genomic DNA was denatured and slotted on the HYBOND™ membrane. D21S5 or human repetitive sequences were used as reference probes for slot blots. The density ratio between the breast cancer gene and the reference gene was calculated for each sample. 10-15 samples of placental DNA digests were used as control. Amongst the control samples, the highest density ratio was set at 1.0. The density ratio of the tumor cell lines were standardized accordingly. An arbitrary cut-off for the standardized ratio (typically 1.3) was defined to identify samples in which the putative gene had been duplicated. Each of the cell lines in the breast cancer panel was scored positively or negatively for duplication of the gene being tested.

Some of the cell lines in the panel were known to have duplicated chromosomal regions from comparative genomic hybridization analysis. In instances where the cDNA being used as probe mapped to the known amplified region, the cDNA indicated that the corresponding gene had also been duplicated. However, duplicated genes were also detected using each of the four cDNAs in instances where comparative genomic hybridization had not revealed any amplification.

Because of the nature of the technique, the standardized ratio calculated as described underestimates the gene copy number, although it is expected to rank in the same order. For example, the standardized ratio obtained for the c-*myc* gene in the SKBR3 breast cancer cell was 5.0. However, it is known that SKBR3 has approximately 50 copies of the c-*myc* gene.

To test for overabundance of RNA, 10 μg of total RNA from breast cancer cell lines or primary breast cancer tumors were electrophoresed on 0.8% agarose in the presence of the denaturant formamide, and then transferred to a nylon membrane. The membrane was probed first with 32P-labeled cDNA corresponding to the putative breast cancer gene, then stripped and reprobed with 32P-labeled cDNA for the beta-actin gene to adjust for differences in sample loading. Ratios of densities between the candidate gene and the beta-actin gene were calculated. RNA from three different cultured normal epithelial cells were included in the analysis as a control for the normal level of gene expression. The highest ratio obtained from the normal cell samples was set at 1.0, and the ratios in the various tumor cells were standardized accordingly.

Example 4: Chromosome 1 gene CH1-9a11-2

One of the cDNA obtained through the selection procedures of Examples 1 and 2 corresponded to a gene that mapped to Chromosome 1.

Table 4 summarizes the results of the analysis for gene duplication and RNA overabundance. Both quantitative and qualitative assessment is shown. The numbers shown were obtained by comparing the autoradiograph intensity of the hybridizing band in each sample with that of the controls. Several control samples were used for the gene duplication experiments, consisting of different preparations of placental DNA. The control sample with the highest level of intensity was used for standardizing the other values. Other sources used for this analysis were breast cancer cell lines with the designations shown. For reasons stated in Example 3, the quantitative number is not a direct indication of the gene copy number, although it is expected to rank in the same order. Similarly, up to 6 control samples were used for the RNA overabundance experiments, consisting of different preparations of breast cell organoids which had been maintained briefly in tissue culture until the experiment was performed. The control sample with the highest level of intensity was used for standardizing the other values. Each cell line was scored + or - according to an arbitrary cut-off value.

TABLE 4: Chromosome 1 Gene In Breast Cancer Cell Lines

Source	CH1-9a11-2 Gene Duplication		CH1-9a11-2 RNA Overabundance 5.2kb		4.4kb	
Normal	-	1.00*	-	1.00**	-	1.0**
BT474	+	2.70	+	1.57	+	3.7
ZR-75-30	+	2.65		nd		nd
MDA453	+	2.86	+	5.79	+	6.2
MDA435	+	3.72	-	0.89	+	2.4
SKBR3	+	1.86	-	0.94	+	2.9
600PE	+	1.72	+	4.47	+	6.8
MDA157	+	1.49	-	1.08	+	1.4
MCF7	+	1.95		nd		nd
DU4475	+	2.02	-	1.13	+	1.5
MDA231	-	1.23	+	1.47	-	
BT20	-	1.09	-	0.83	+	1.9
T47D	-	1.05		nd		nd
UACC812	-	0.67	+	1.57	+	1.8
MDA134	-	1.19	+	5.04	+	7.1
CAMA-1	-	1.02	+	2.51	+	7.2
Incidence (%)	**9/15 (60%)**		**7/12 (58%)**		**11/12 (92%)**	

Gene duplication or RNA overabundance; - no duplication or overabundance; nd = not done

* Degree of gene duplication is reported relative to placental DNA preparations.

** Degree of RNA overabundance is reported relative to the highest level observed for several cultures of normal epithelial cells. Two hybridizing species of RNA are calculated and reported separately.

The gene corresponding to the CH1-9a11-2 cDNA was duplicated in 9 out of 15 (60%) of the breast cancer cell lines tested, compared with placental DNA digests (P3 and P12). The sequence of the 115 bases from the 5' end of the cDNA fragment (SEQ. ID NO:1) is shown in Figure 22. There was no substantial homology to any known gene in GenBank. One of the three possible reading frames was found to be open, with the predicted amino acid shown in Figure 22 (SEQ. ID NO:2).

The CH1-9a11-2 gene was further characterized by obtaining additional sequence information. A λ-GT10 cDNA library from the breast cancer cell line BT474 (Example 2) was screened using the initial cDNA insert, and a clone with a 2.5 kilobase insert was identified. The identified clone was subcloned into plasmid vector pCRII. T7 and Sp6 primers for regions flanking the cDNA inserts were used as initial sequencing primers:

T7 primer: (SEQ. ID NO:42)

5'-TAATACGACTCACTATAGGGAGA-3'

Sp6 primer: (SEQ. ID NO:43)

5'-CATACGATTTAGGTGACACTATAG-3'

Sequencing continued by walking along the region of interest by standard techniques, using sequencing primers based on data already obtained. Primers used in sequencing are designated 1-16 in Figure 7.

A second clone (designated pCH1-1.1) overlapping on the 5' end was obtained using CLONTECH Marathon™ cDNA Amplification Kit. A map showing the overlapping regions is provided in Figure 6. Briefly, two DNA primers designated CH1a and CH1b (Figure 7) were synthesized. Polyadenylated RNA from breast cancer cell line 600PE was reverse transcribed using CH1b primer. After second strand synthesis, adaptor DNA provided in the kit was ligated to the double-stranded cDNA. The 5' end cDNA of CH1-9a11-2 was then amplified by PCR using primers CH1a and AP1 (provided in the kit). To increase the specificity of the PCR products, the first PCR products were PCR reamplified using nested primers CH1a and AP2 (provided in the kit). The PCR products were cloned into pCRII vector (Invitrogen) and screened with CH1-9a11-2 probe.

The sequence of 3452 base pairs between the 5' end of pCH1-1.1 and the poly-A tail of CH1-9a11-2 was determined by standard sequencing techniques. The DNA sequence is shown in Figure 8 (SEQ. ID NO:15). The longest open reading frame is in frame 1 (bases 1-1875), and codes for 624 amino acids before the stop codon. The corresponding amino acid sequence of this frame is shown in the upper panel of Figure 9 (SEQ. ID NO:16). The partial sequence predicted for the translated protein is listed the low panel of Figure 9 (SEQ. ID NO:17). Bases 1876 to the end of the sequence are believed to be a 3' untranslated region. A hydrophobicity analysis identified a putative membrane insertion or membrane spanning region at about amino acids 382-400, indicated in Figure 9 by underlining.

Figure 23 is a listing of additional cDNA sequence obtained for CH1-9a11-2, comprising approximately 1934 base pairs 5' from the sequence of Figure 8. The additional sequence data was obtained by rescuing and amplifying two further fragments of CH1-9a11-2 cDNA. Nested primers were designed ~100 base pairs downstream from the 5' end of the known sequence. The primers were used in a nested amplification assay using AP1 and AP2, using the CLONTECH Marathon™ cDNA Amplification Kit as described above. The template for the first upstream fragment was reverse-transcribed polyadenylated RNA from breast cancer cell line 600PE , as described earlier.

This fragment was sequenced, and another set of nested primers was designed. The template for the next upstream fragment was a Marathon™ ready cDNA preparation from human testes, also supplied by CLONTECH.

The nucleotide sequence shown in Figure 23 comprises an open reading frame through to the 5' end. Figure 24 shows the corresponding protein translation. Between about another 500-1000 bases are predicted to be present in the CH1-9a11-2 direction, with the protein encoding sequence beginning somewhere within this additional sequence. Sequencing of the encoding region is completed by obtaining additional CH1-9a11-2 fragments in this direction.

A GENINFO® BLAST search of nucleotide and peptide sequence databases was performed through the National Center for Biotechnology Information on February 23, 1996. Short segments of homology with other reported human sequences were found at the nucleotide level (<500 base pairs), but none with any ascribed function in the respective identifier. At the amino acid level, no identity higher than 30% was found with any reported eukaryotic sequences.

A CH1-9a11-2 cloned insert has been used to probe the level of relative expression in polyadenylated RNA from a panel of tissue sources. The RNA was obtained already prepared for Northern blot analysis (CLONTECH Catalog # 7759-1, 7760-1 and 7756-1.) The manufacturer produced the blots from approximately 2 µg of poly-A RNA per lane, run on a denaturing formaldehyde 1-2% agarose gel, transferred to a nylon membrane, and fixed by UV irradiation. The relative CH1-9a11-2 expression observed at the RNA level is shown in Table 5:

TABLE 5: Northern blot analysis

Tissue	CH1-9a11-2 mRNA
heart	++
brain	+
placenta	++
lung	+/–
liver	+/–
skeletal muscle	+
kidney	+/–
pancreas	+++
spleen	+
thymus	+
prostate	++
testis	+++
ovary	++
small intestine	+
colon	+/–
peripheral blood	+/–

++++ Very high
+++ High
++ Medium
+ Low
+/– Very low

Relatively elevated levels of expression were observed in heart, placenta, pancreas, prostate, testis and ovary. The level of expression in breast cancer cell lines is also relatively high (about ++++ on the scale), since the Northern analysis performed on these lines (described above) was conducted on *total* cellular RNA, of which polyadenylated RNA constitutes only about 5%. It is likely that the CH1-9a11-2 gene is involved in a biological process that is typical to the tissue types showing medium to high levels of expression, which may relate to increased tissue growth or metabolism.

Since the obtained sequence is shorter than the apparent size of mRNA observed in Northern analysis (Table 1), an additional polynucleotide segment is believed to be present at the 5' end of the sequence shown in SEQ. ID NO:15. Further sequence data at the 5' end is deduced by obtaining additional cloned cDNA using standard techniques. Briefly, in one approach, mRNA from breast cancer cell lines MDA-453 and/or 600PE are cloned and screened using primers based on sequence data from SEQ. ID NO:15. Two nested primers of about 20 nucleotides are prepared, the innermost about 150 base pairs from the 5' end, and the outermost about 170 base pairs from the 5' end. The outermost primer is used to synthesize a first cDNA strand complementary to the mRNA in the upstream direction. Second strand synthesis is performed using reagents in a CLONTECH

Marathon™ cDNA amplification kit according to manufacturer's directions. The double-stranded DNA is then ligated at the 5' end of the coding sequence with the double-stranded adaptor fragment provided in the kit. A first PCR amplification (about 30 cycles) is performed using the first adapter primer from the kit and the outermost RNA-specific primer, and a second amplification (about 30 cycles) is performed using the second adapter primer and the innermost RNA-specific primer. In an alternative approach, a CLONTECH RACE-READY single-stranded cDNA from human placenta is PCR amplified using nested 5' anchor primers in combination with the outermost and innermost RNA-specific primers. Amplified DNA obtained using either approach is analyzed by gel electrophoresis, and cloned into plasmid vector pCRII. Clones are screened, as necessary, using the 2.5 kilobase CH1-9a11-2 insert. Clones corresponding to full-length mRNA (4.5 kb or 5.5 kb; Table 1), or cDNA fragments overlapping at the 5' end are selected for sequencing. Compared with the 4.5 kb form, additional polynucleotide segments may be present in the 5.5 kb form within the encoding region, or in the 5' or 3' untranslated region.

Example 5: Chromosome 8 gene CH8-2a13-1

One of the cDNA obtained corresponded to a gene that mapped to Chromosome 8. Figure 2 shows the Southern blot analysis for the corresponding gene in various DNA digests. Lane 1 (P12) is the control preparation of placental DNA; the rest show DNA obtained from human breast cancer cell lines. Panel A shows the pattern obtained using the 32P-labeled CH8-2a13-1 cDNA probe. Panel B shows the pattern obtained with the same blot using the 32P-labeled D2S6 probe as a loading control. The sizes of the restriction fragments are indicated on the right.

Figure 3 shows the Northern blot analysis for RNA overabundance. Lanes 1-3 show the level of expression in cultured normal epithelial cells. Lanes 4-19 show the level of expression in human breast cancer cell lines. Panel A shows the pattern obtained using the CH8-2a13-1 probe; panel B shows the pattern obtained with beta-actin cDNA, a loading control.

The results are summarized in Table 6. The scoring method is the same as for Example 4.

TABLE 6: Chromosome 8 Genes In Breast Cancer Cell Lines

Source	CH8-2a13-1 Gene Duplication		CH8-2a13-1 RNA Overabundance		*c-myc* Gene Duplication	
Normal	-	1.00*	-	1.00**	-	1.00*
SKBR3	+	4.25	+	4.30	+	4.73
ZR-75-30	+	3.82	nd		+	2.24
BT474	+	1.53	+	1.72	+	1.76
MDA157	+	2.02	+	3.39	+	1.39
MCF7	+	1.84	+	4.92	+	3.10
CAMA-1	+	3.62	+	2.14	+	1.61
MDA361	+	2.00	+	1.74	nd	
MDA468	nd		+	4.50	nd	
T47D	+	1.41	+	1.58	-	1.02
MDA453	+	1.83	+	3.10	-	0.90
MDA134	+	1.30	+	3.70	-	0.88
MDA435	+	2.15	+	4.94	-	1.00
600PE	-	0.95	+	2.04	-	0.54
UACC812	+	1.25	+	2.40	-	0.74
MDA231	-	0.80	+	1.28	+	1.27
DU4475	-	0.85	-	0.88	-	0.50
BT468	-	0.37	-	0.70	-	0.23
BT20	-	0.95	-	0.82	-	
Incidence (%)	**12/17 (71%)**		**14/17 (82%)**		**7/16 (44%)**	

\+ Gene duplication or RNA overabundance; - no duplication or overabundance; nd = not done.
* Degree of gene duplication is reported relative to placental DNA preparations.
** Degree of RNA overabundance is reported relative to the highest level observed for several cultures of normal epithelial cells.

The gene corresponding to CH8-2a13-1 showed clear evidence of duplication in 12 out of 17 (71%) of the cells tested. RNA overabundance was observed in 14 out of 17 (82%). Thus, 11% of the cells had achieved RNA overabundance by a mechanism other than gene duplication.

Since the known oncogene c-myc is located on Chromosome 8, the Southern analysis was also conducted using a probe for c-myc. At least 2 of the breast cancer cells showing duplication of the gene corresponding to CH8-2a13-1 gene did not show duplication of c-myc. This indicates that the gene corresponding to CH8-2a13-1 is not part of the myc amplicon.

The sequence of 150 bases from the 5' end of the cDNA fragment is shown in Figure 22 (SEQ ID NO:3). There was no substantial homology to any known gene in GenBank. One of the

three possible reading frames was found to be open, with the amino acid sequence shown in Figure 22 (SEQ ID NO:4).

The CH8-2a13-1 gene was further characterized by obtaining additional sequence information. A λ-GT10 cDNA library from the breast cancer cell line BT474 (Example 2) was screened using the initial cDNA insert, and clones with a 3.0 kb and a 4.0 kb insert were identified. The two identified clones were subcloned into plasmid vector pCRII. T7 and Sp6 primers for regions flanking the cDNA inserts were used as initial sequencing primers. Sequencing continued by walking along the region of interest by standard techniques, using sequencing primers based on data already obtained. The two inserts were found to overlap (Figure 6). Primers used are those designated 1-25 in Figure 10.

A third clone of about 600 bp (designated pCH8-600) overlapping on the 5' end (Figure 6) was obtained using CLONTECH Marathon™ cDNA Amplification Kit. Briefly, two DNA primers CH8a and CH8b (Figure 10) were synthesized. Polyadenylated RNA from breast cancer cell line BT474 was reverse transcribed using CH8b primer. After second strand synthesis, adaptor DNA provided in the kit was ligated to the double-stranded cDNA. The 5' end cDNA of CH8-2a13-1 was then amplified by PCR using primers CH8a and AP1 (provided in the kit). To increase the specificity of the PCR products, the first PCR products were PCR reamplified using nested primers CH8a and AP2 (provided in the kit). The PCR products were cloned into pCRII vector (Invitrogen) and screened with CH8-2a13-1 probe.

By sequencing relevant portions of the three clones, a nucleic acid sequence of 3982 base pairs between the 5' end and the poly-A tail of CH8-2a13-1 was determined. The DNA sequence is shown in Figure 11 (SEQ. ID NO:18). Bases 1-152 are believed to be a 5' untranslated region. The longest open reading frame is in frame 3 from base 153 to 3911, and codes for 1252 amino acids before the stop codon. The corresponding amino acid sequence of this frame is shown in the upper panel of Figure 12 (SEQ. ID NO:19). The sequence predicted for the translated protein is shown in the lower panel of Figure 12(SEQ. ID NO:20).

A GENINFO® BLAST search of nucleotide and peptide sequence databases was performed through the National Center for Biotechnology Information on March 26, 1996. The sequences were found to be about 99% identical at the nucleotide and amino acid level with bases 343-4103 of KIAA0196 protein (N. Nomura et al., in press; sequence submitted to the DDBJ/EMBL/GenBank databases on March 4, 1996). The KIAA0196 was one of 200 different cDNA cloned at random from an immature male human myeloblast cell line. KIAA0196 has no known biological function, and is described by Nomura et al. as being ubiquitously expressed.

A fourth clone of about 600 bp overlapping pCH8-600 at the 5' end has also been obtained. Briefly, a DNA primer was synthesized corresponding to about the first 20 nucleotides at the 5' of the predicted cDNA sequence, and used along with a primer based on the pCH8-600 sequence to reverse-transcribe RNA from breast cancer cell line BT474. The product was cloned into pCRII vector (Invitrogen) and screened with a CH8-2a13-1 probe. The new clone is sequenced along both strands to obtain additional 5' untranslated sequence data for the cDNA. The predicted compiled cDNA

nucleotide sequence of CH8-2a13-1 cDNA is shown in Figure 13 (SEQ. ID NO:21). The corresponding amino acid sequence of this frame is shown in Figure 14 (SEQ. ID NO:22). A polynucleotide comprising the compiled sequence is assembled by joining the insert of this fourth clone to pCH8-4k within the shared region. Briefly, CH8-4k is cut with *Xba*I and *Not*I. The fourth clone is cut with *Bam*HI and *Xba*I. The ligated polynucleotide is then inserted into pCRII cut with *Bam*HI and *Not*I.

A CH8-2a13-1 cloned insert has been used to probe the level of relative expression in polyadenylated RNA from a panel of tissue sources obtained from CLONTECH, as in Example 4. The relative CH8-2a13-12 expression observed at the mRNA level is shown in Table 7:

TABLE 7: Northern blot analysis

Tissue	CH1-9a11-2 mRNA
heart	++
brain	+
placenta	+
lung	+
liver	+/–
skeletal muscle	+/–
kidney	+/–
pancreas	+/–
spleen	+
thymus	+
prostate	+
testis	++
ovary	+
small intestine	+
colon	+
peripheral blood	+/–

++++ Very high
+++ High
++ Medium
+ Low
+/– Very low

Relative levels of expression observed were as follows: Low levels of expression were observed in adult peripheral blood leukocytes (PBL), brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. Medium levels of expression were observed in adult heart, spleen, thymus, prostate, testis, ovary, small intestine, and colon. High levels of expression were observed in four fetal tissues tested: brain, lung, liver and kidney. The level of expression in breast cancer cell lines is relatively high

(about ++++ on the scale), since the Northern analysis performed on these lines was conducted on *total* cellular RNA. It is likely that the CH8-2a13-1 gene is involved in a biological process that is typical to the tissue types showing medium to high levels of expression, which may relate to increased tissue growth or metabolism.

Example 6: Chromosome 13 gene CH13-2a12-1

One of the cDNA obtained corresponded to a gene that mapped to Chromosome 13. Figure 4 shows the Southern blot analysis for the corresponding gene in various DNA digests. Lanes 1 and 2 are control preparations of placental DNA; the rest show DNA obtained from human breast cancer cell lines. Panel A shows the pattern obtained using the CH13-2a12-1 cDNA probe; panel B shows the pattern using D2S6 probe as a loading control. The sizes of the restriction fragments are indicated on the right.

Figure 5 shows the Northern blot analysis for RNA overabundance of the CH13-2a12-1 gene. Lanes 1-3 show the level of expression in cultured normal epithelial cells. Lanes 4-19 show the level of expression in human breast cancer cell lines. Panel A shows the pattern obtained using the CH13-2a12-1 probe; panel B shows the pattern obtained with beta-actin cDNA, a loading control. The apparent size of the mRNA varied depending upon conditions of electrophoresis. Full-length mRNA is believed to occur at sizes of about 3.2 and 3.5 kb.

The results of the RNA abundance comparison are summarized in Table 8. The scoring method is the same as for Example 4.

TABLE 8: Chromosome 13 Gene In Breast Cancer Cell Lines

Source	CH13-2a12-1 Gene duplication		CH13-2a12-1 RNA Overabundance	
Normal	-	1.00*	-	1.00**
600PE	+	2.18	+	5.57
BT474	+	1.60	+	3.20
SKBR3	+	1.58	+	4.25
MDA157	+	2.21	+	3.76
CAMA-1	+	1.41	+	1.99
MDA231	+	1.65	+	2.09
T47D	+	1.23	+	1.20
MDA468	nd		+	6.90
MDA361	nd		+	2.59
MDA435	-	0.59	+	3.41
MDA134	-	0.53	+	2.59
DU4475	-	0.75	+	1.79
MDA453	-	0.89	+	1.97
BT20	-	0.37	-	1.04
MCF7	-	0.29	-	1.03
UACC812	-	0.30	-	0.39
BT468	-	0.47	nd	
ZR-75-30	-	0.70	nd	
Incidence (%)	**7/16 (44%)**		**13/16 (81%)**	

\+ Gene duplication or RNA overabundance; - no duplication or overabundance; nd = not done
* Degree of gene duplication is reported relative to placental DNA preparations.
** Degree of RNA overabundance is reported relative to the highest level observed for several cultures of normal epithelial cells.

The gene corresponding to CH13-2a12-1 was duplicated in 7 out of 16 (44%) of the cells tested. Three of the positive cell lines (600PE, BT474, and MDA435) had been studied previously by comparative genomic hybridization, but had not shown amplified chromatin in the region where CH13-2A12-1 has been mapped in these studies.

RNA overabundance was observed in 13 out of 16 (81%) of the cell lines tested. Thus, 37% of the cells had achieved RNA overabundance by a mechanism other than gene duplication.

Cells from primary breast tumors have also been analyzed them for duplication of the chromosome 13 gene. Ten of the 82 tumors analyzed (12%) were positive, confirming that duplication of this gene is not an artifact of in vitro culture.

The sequence of 107 bases from the 5' end of the 1.5 kb cDNA fragment is shown in Figure 22 (SEQ ID NO:5). There was no substantial homology to any known gene in GenBank. One of the three possible reading frames was found to be open, with the predicted amino acid sequence shown in Figure 22 (SEQ ID NO:6).

The CH13-2a12-1 gene was further characterized by obtaining additional sequence information. A λ-GT10 cDNA library from the breast cancer cell line BT474 (Example 2) was screened using the initial cDNA insert, and clones with a 3.5 kilobase and a 1.6 kilobase insert were identified. The two identified clones were subcloned into plasmid vector pCRII. T7 and Sp6 primers for regions flanking the cDNA inserts were used as initial sequencing primers. Sequencing continued by walking along the region of interest by standard techniques, using sequencing primers based on data already obtained. The two inserts were found to overlap (Figure 6). Primers used during sequencing are shown in Figure 15.

By sequencing relevant portions of the 3.5 and 1.6 kb clones, a nucleic acid sequence of 3339 base pairs between the 5' end and the poly-A tail of CH13-2a12-1 was determined. The DNA sequence is shown in Figure 16 (SEQ. ID NO:23). Bases 1-520 are believed to be a 5' untranslated region. The longest open reading frame is in frame 2 from base 521 to 1838, and codes for 611 amino acids before the stop codon. The corresponding amino acid sequence of this frame is shown in the upper panel of Figure 17 (SEQ. ID NO:24). The sequence predicted for the translated protein is shown in the lower panel of Figure 17 (SEQ. ID NO:25). Bases 1838 to 3339 of the nucleotide sequence are believed to be a 3' untranslated region, which is present in the 3.5 kb insert. The 3.5 kb insert appears to be a splice variant (Figure 6), in which the 3' untranslated region consists of bases 1838-2797 in the sequence.

A GENINFO® BLAST search of nucleotide and peptide sequence databases was performed through the National Center for Biotechnology Information on March 26, 1996. Short segments of homology with other reported human sequences were found at the nucleotide level (<500 base pairs), but none with any ascribed function in the respective identifier. At the amino acid level, the sequence was found to share 33% identities and 54% positives with 228 residues of the *lin 19* protein of *Caenorhabditis elegans*. This protein has been implicated in regulating the cell cycle of *C. elegans* (ET Kiprecs, W He & EM Hedgecock). The CH13-2a12-1 gene is suspected of a role in controlling cell proliferation. "Controlling cell proliferation" in this context means that an abnormally high or low level of gene expression at the RNA or protein level results in a higher or lower rate of cell proliferation, or vice versa, compared with cells with an otherwise similar phenotype. There is also a low-level homology between CH13-2a12-1 and VACM-1, a vasopressin-activated, calcium-mobilizing receptor from rabbit kidney medulla (Burnatowska-Hledin et al). VACM-1 has a transmembrane

sequence, whereas none has been detected in CH13-2a12-1. Nevertheless, it is possible that the CH13-2a12-1 protein product has a Ca^{++} binding or Ca^{++} mobilizing function.

A CH13-2a12-1 cloned insert has been used to probe the level of relative expression in polyadenylated RNA from a panel of tissue sources obtained from CLONTECH, as in Example 4. The relative CH13-2a12-1 expression observed at the mRNA level is shown in Table 9:

TABLE 9: Northern blot analysis

Tissue	CH13-2a12-1 mRNA
heart	++++
brain	+
placenta	++
lung	+
liver	++
skeletal muscle	++++
kidney	+
pancreas	++
spleen	++
thymus	++
prostate	++
testis	+++
ovary	++
small intestine	++
colon	+
peripheral blood	+

++++ Very high
+++ High
++ Medium
+ Low
+/– Very low

Relatively elevated levels of expression were observed in heart, skeletal muscle and testis. The level of expression in breast cancer cell lines is relatively high (about ++++ on the scale), since the Northern analysis performed on these lines was conducted on *total* cellular RNA. It is likely that the CH13-2a12-1 gene is involved in a biological process that is typical to the tissue types showing medium to high levels of expression, which may relate to increased tissue growth or metabolism.

Fragments corresponding to the CH13-2a12-1 gene have also been used to screen cell lines derived from other types of cancer. Southern analysis showed that about 1 out of 4 breast cancer cell lines tested have gene duplication of CH13-2a12-1. Northern analysis showed that about 3 out of 6 lines tested have overexpression of the corresponding RNA transcript.

Example 7: Chromosome 14 gene CH14-2a16-1

One of the cDNA obtained corresponded to a gene that mapped to Chromosome 14. Results of the analysis are summarized in Table 10. The scoring method is the same as for Example 4.

TABLE 10: Chromosome 14 Gene In Breast Cancer Cell Lines

Source	CH14-2a16-1 Gene duplication		CH14-2a16-1 RNA Overabundance	
Normal	-	1.00*	-	1.00**
BT474	+	2.89	+	2.57
MCF7	+	1.35	+	1.88
SKBR3	+	2.58	+	2.19
T47D	+	2.28	nd	
MDA157	+	1.52	+	2.52
UACC812	+	2.23	nd	
MDA361	-	0.97	+	1.43
MDA453	+	1.58	+	5.92
BT20	-		-	1.07
600PE	-	0.94	+	2.00
MDA231	+	1.66	+	2.19
CAMA-1	-	0.92	-	0.71
DU4475	-	0.87	+	1.33
BT468	-	0.46	nd	
MDA134	-	0.77	+	7.17
Incidence (%)	**8/15 (53%)**		**10/12 (83%)**	

\+ Gene duplication or overabundance; - no duplication or overabundance; nd = not done
* Degree of gene duplication is reported relative to placental DNA preparations.
** Degree of RNA overabundance is reported relative to the highest level observed for several cultures of normal epithelial cells.

The gene corresponding to CH14-2a16-1 was duplicated in 8 out of 15 (53%) of the cells tested. The sequence of 114 bases from the 5' end of the cDNA fragment is shown in Figure 22 (SEQ ID NO:7). There was no substantial homology to any known gene in GenBank. One of the three possible reading frames was found to be open, with the predicted amino acid sequence shown in Figure 22 (SEQ ID NO:8).

The CH14-2a16-1 gene was further characterized by obtaining additional sequence information. A λ-GT10 cDNA library from the breast cancer cell line BT474 (Example 2) was screened using the initial cDNA insert, and two clones were identified: one with a 1.6 kb insert, and the other with a 2.5 kb insert. The identified clones were subcloned into plasmid vector pCRII. The 1.6 kb insert was sequenced by using T7 and Sp6 primers for regions flanking the cDNA inserts as initial sequencing primers. Sequencing continued by walking along the region of interest by standard techniques, using sequencing primers based on data already obtained. Primers used are those designated 1-11 in Figure 18.

A third clone (designated pCH14-800) overlapping on the 5' end (Figure 6) was obtained using CLONTECH Marathon™ cDNA Amplification Kit. Briefly, DNA primers CH14a, CH14b, CH14c and CH14d (Figure 18) were prepared. Polyadenylated RNA from breast cancer cell line MDA453 was reverse transcribed using 14b primer. After second strand synthesis, adaptor DNA provided in the kit was ligated to the double-stranded cDNA. The 5' end cDNA of CH14-2a16-1 was then amplified by PCR using primers CH14b (or CH14c) and AP1 (provided in the kit). To increase the specificity of the PCR products, the first PCR products were PCR reamplified using nested primers CH14a (or CH14d) and AP2 (provided in the kit). The PCR products were cloned into pCRII vector (Invitrogen) and screened with CH14-2a16-1 probe.

By sequencing pCH14-1.6 and pCH14-800, a nucleic acid sequence of 2021 base pairs between the 5' end and the poly-A tail of CH14-2a16-1 has been determined. The DNA sequence is shown in Figure 19 (SEQ. ID NO:26). The longest open reading frame is in frame 1 from base 1 to 792, and codes for 263 amino acids before the stop codon. The corresponding amino acid sequence of this frame is shown in the upper panel of Figure 20 (SEQ. ID NO:27). The partial sequence predicted for the translated protein is shown in the lower panel of Figure 20 (SEQ. ID NO:28). The 2.1 kb clone has not been sequenced, but is believed to consist about the same region of the CH14-2a16-1 cDNA as pCH14-1.6 and pCH14-800 combined.

A GENINFO® BLAST search of nucleotide and peptide sequence databases was performed through the National Center for Biotechnology Information on March 26, 1996. Short segments of homology with other reported human sequences were found at the nucleotide level (<500 base pairs), but none with any ascribed function in the respective identifier. At the amino acid level, the sequence was found to share homologies within the first 106 residues with an RNA binding protein from *Saccharomyces cerevisiae* with the designation *NAB2*. NAB2 is one of the major proteins associated with nuclear polyadenylated RNA in yeast cells, as detected by UV light-induced cross-linking and immunofluorescence. NAB2 is strongly and specifically associated with nuclear poly(A)+ RNA in vivo. Gene knock-out experiments have shown that this protein is essential to yeast cell survival (Anderson et al.). Accordingly, the protein encoded by CH14-2a16-1 is suspected of having DNA or RNA binding activity.

A fourth clone (pCH14-1.3) has been obtained that overlaps the pCH14-800 clone at the 5' end (Figure 6). The method of isolation was similar to that for pCH14-800, using primers based on the pCH14-800 sequence. Partial sequence data for pCH14-1.3 has been obtained by one-

directional sequencing from the 5' and 3' ends of the pCH14-1.3 clone. Figure 21 shows the nucleotide sequence of the sequence of the 5' end (SEQ. ID NO:29) and the amino acid translation of the likely open reading frame (SEQ. ID NO:30); the nucleotide sequence of the 3' end (SEQ. ID NO:31) and the likely open reading frame (SEQ. ID NO:32). This data is confirmed and additional sequence between SEQ. ID NOS.29 and 31 is obtained by fully sequencing both strands of pCH14-1.3. Once compiled, the sequence data from pCH14-1.3, pCH14-800 and pCH14-1.6 may be shorter than the apparent size of mRNA observed in Northern analysis (Table 1). If necessary, further sequence data at the 5' end is deduced by obtaining additional cloned cDNA according to approaches described in this Example or Example 4.

Figure 25 is a listing of additional cDNA sequence obtained for CH14-2a16-1, comprising approximately 1934 base pairs 5' from the sequence of Figure 19. The corresponding amino acid translation is shown in the upper panel of Figure 26. The additional sequence data was obtained by rescuing and amplifying further fragments of CH14-2a16-1 cDNA. Nested primers were designed ~100 base pairs downstream from the 5' end of the known sequence. The primers were used in a nested amplification assay using AP1 and AP2, using the CLONTECH Marathon™ cDNA Amplification Kit as described above. The template was a Marathon™ ready cDNA preparation from human testes, also supplied by CLONTECH.

The nucleotide sequence shown in Figure 25 is closed at the the 5' end. The lower panel of Figure 26 shows what is predicted to be the sequence of the gene product, beginning at the first methionine residue. The nucleotide sequence shown contains a point difference at the position indicated by the underlining in Figure 25. A base determined to be A from the previously obtained polynucleotide fragment was a G in the one used in this part of the experiment. This corresponds to a change from E (glutamic acid) to G (glycine) in the protein sequence, at the position underlined in Figure 26. This may represent a natural allelic variation.

A CH14-2a16-1 cloned insert has been used to probe the level of relative expression in polyadenylated RNA from a panel of tissue sources obtained from CLONTECH, as in Example 4. The relative CH14-2a16-1 expression observed at the mRNA level is shown in Table 11:

TABLE 11: Northern blot analysis

Tissue	CH14-2a16-1 mRNA
heart	+
brain	+
placenta	+
lung	+
liver	+
skeletal muscle	+
kidney	+/–
pancreas	+
spleen	+
thymus	+
prostate	+
testis	++++
ovary	+
small intestine	+
colon	+
peripheral blood	+/

++++ Very high
+++ High
++ Medium
+ Low
+/– Very low

CH14-2a16-1 mRNA was particularly high in testis. The level of expression in breast cancer cell lines is also quite high, since the Northern analysis performed on these lines was conducted on *total* cellular RNA. It is likely that the CH14-2a16-1 gene is involved in a biological process that is typical to the tissue types showing medium to high levels of expression, which may relate to increased tissue growth or metabolism.

Five motifs corresponding to a zinc finger protein have been found in the CH14-2a16-1 nucleotide sequence. Further zinc finger motifs may be present in CH14-2a16-1 in the upstream direction. Zinc finger motifs are present, for example, in RNA polymerases I, II, and III from *S. cerevisiae*, and are related to the zinc knuckle family of RNA/ssDNA-binding proteins found in the HIV nucleocapsid protein. The actual sequence observed in each of the five zinc finger motifs of CH14-2a16-1 is:

Cys–(Xaa)$_5$–Cys–(Xaa)$_4$–Cys–(Xaa)$_3$–His or (SEQ. ID NO:38)

Cys–(Xaa)$_5$–Cys–(Xaa)$_5$–Cys–(Xaa)$_3$–His (SEQ. ID NO:39)

which is indicated in Figure 20 by underlining. This is identical to the 7 zinc finger motifs of NAB2, which make up an RNA/ssDNA binding region (Anderson et al.). Accordingly, the CH14-2a16-1 gene product is suspected of having DNA or RNA binding activity, and may be specific for polyadenylated RNA. It may very well play a role in the regulation of gene replication, transcription, the processing of hnRNA into mature mRNA, the export of mRNA from the nucleus to the cytoplasm, or translation into protein. This role in turn may be closely implicated in cell growth or proliferation, particularly as manifest in tumor cells.

Example 8: Identification of other cancer-associated genes

cDNA fragments corresponding to additional cancer-associated genes are obtained by applying the techniques of Examples 1 & 2 with appropriate adaptations. As before, cancer cells are selected for use in differential display of RNA, based on whether they share a duplicated chromosomal region according to Table 12:

TABLE 12: Cancer cell lines sharing duplicated chromosomal regions

Chromosomal location	Cancer type & references
1p22-32	small cell (Levin 1994)
1p22	bladder (Kallioniemi 1995)
1p32-33	rabdomyosarcoma (Steilen-Gimbel); breast (Ried 1995); small cell lung (Ried 1994)
1q21-22	sarcoma (Forus 1995a & b); breast (Muleris 1994a)
1q24	small cell (Levin 1994)
1q31	bladder (Kallioniemi 1995)
1q32	glioma (Muleris 1994b; Schrock)
1q	head and neck (Speicher 1995), breast (Muleris 1994a)
2p23	small cell lung (Ried 1994)
2p24-25	small cell lung (Levin 1994)
2	head and neck (Speicher 1995)
2q	head and neck (Speicher 1995)
2q33-36	head and neck (Speicher 1995)
3p22-24	bladder (Voorter), small cell (Levin 1994)
3q24-26	bladder (Kallioniemi 1995), glioma (Kim), osteosarcoma (Tarkkanen)
3q25-26	ovarian (Iwabuchi)

TABLE 12: Cancer cell lines sharing duplicated chromosomal regions

Chromosomal location	Cancer type & references
3q26-term	head and neck (Speicher 1995)
3q	small cell lung (Levin 1995; Rerid 1994); head and neck (Speicher 1995)
4q12	glioma (Schrock)
5p	small cell lung (Levin 1994 & 1995; Ried 1994)
5p15.1	glioma (Muleris 1994b)
6p	osteosarcoma (Forus 1995a); breast (Ried 1995)
6p21-term	melanoma (Speicher)
7p	glioma (Schliegel 1994 & 1996; may be EGFR)
7p11-12	glioma (Muleris 1994b; Schrock), small cell lung (Ried 1994)
7q21-32	glioma (Kim; Muleris 1994b; Schrock)
7q21-22	head and neck (Speicher), glioma (Schrock)
7q33-term	head and neck (Speicher 1995)
7	colon (Schlegel 1995); glioma (Kim), head and neck (Speicher); prostate (Visakorpi)
8q	small cell lung (Ried 1994)
8q21	bladder (Kallioniemi 1995)
8q24	myeloid leukemia (Mohamed)
8q22-24	glioma (Kim; Muleris 1994b); breast (Muleris 1994a)
8q24-25	small cell (Levin 1994; Ried 1994); breast (Muleris 1994a)
8q23-term	sarcoma (Forus 1995a), melanoma (Speicher)
8q24	ovarian (Iwabuchi)
8q	breast (Ried 1995; Isola; Muleris 1994a), small cell lung (Levin 1994 & 1995), B-cell leukemias (Bentz 1994a), myeloid leukemia (Bentz 1994b), glioma (Schlegel), head and neck (Speicher 1995), prostate (Cher, Visakorpi)
9	head and neck (Speicher)
9p	head and neck (Speicher)
9p2	glioma (Muleris 1994b)
9p13	breast (Muleris 1994a)
10p	head and neck (Speicher 1995)
10p13-14	bladder (Voorter)
10q22	breast (Muleris 1994a)
11q13	head and neck (Speicher 1995), breast (Muleris 1994a)

TABLE 12: Cancer cell lines sharing duplicated chromosomal regions

Chromosomal location	Cancer type & references
12	B-cell leukemias (Bentz 1995a)
12p	head and neck (Speicher 1995), glioma (Schrock)
12q	glioma (Schlegel 1994)
12q12-15	bladder (Voorter), osteosarcoma (Tarkkanen), liposarcoma (Suijkerbuijk)
12q21.3-22	liposarcoma (Suijkerbuijk)
13	colon (Schlegel 1995)
13q	breast (Ried 1995), head and neck (Speicher 1995)
13q21-34	bladder (Kallioniemi 1995)
13q32-term	head and neck (Speicher 1995), small cell lung (Ried 1994)
14q	head and neck (Speicher 1995)
15q26	breast (Muleris 1994a)
16	head and neck (Speicher 1995)
16p	breast (Ried 1995)
16p11.2	breast (Muleris 1994a)
17	head and neck (Speicher 1995)
17p11-12	osteosarcoma (Forus 1995a; Tarkkanen)
17q	breast (Ried 1995), small cell lung (Ried 1994)
17q21.1	breast (Muleris 1994a)
17q22-23	bladder (Voorter), breast (Muleris 1994a)
17q22-24	breast (Kallioniemi 1994)
18p11	bladder (Voorter)
19q13.1	small cell lung (Ried 1994)
20p	head and neck (Speicher 1995)
20q	ovarian (Iwabuchi), colon (Schlegel 1995), breast (Isola; Tanner)
20q13.3	breast (Muleris 1994a), Kallioniemi (1994)
22q	head and neck (Speicher 1995)
22q11-13	bladder (Voorter), glioma (Schrock)
X	prostate (Visakorpi)
Xq	small cell lung (Levin 1995)
Xq24	small cell (Levin 1994)
Xq11-13	prostate (Visakorpi), osteosarcoma (Tarkkanen)

Control RNA is prepared from normal tissues to match that of the cancer cells in the experiment. Normal tissue is obtained from autopsy, biopsy, or surgical resection. Absence of neoplastic cells in the control tissue is confirmed, if necessary, by standard histological techniques.

5 cDNA corresponding to RNA that is overabundant in cancer cells and duplicated in a proportion of

the same cells is characterized further, as in Examples 3-7. Additional cDNA comprising an entire protein-product encoding region is rescued or selected according to standard molecular biology techniques as described elsewhere in this disclosure.

REFERENCES

Articles on general topics

1. Adnane J. et al. (1991), "BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers", Oncogene 6:659-661.
2. Alitalo K. et al. (1986), "Oncogene amplification in tumor cells", Adv. Cancer Res. 47:235-281.
3. Altschul et al. (1986), Bull. Math. Bio. 48:603-616.
4. Beardsley T. (1994), "Crabshoot: manufacturers gamble on cancer vaccines again", Scientific American, Sept:102.
5. Berns E.M. et al. (1992), "Sporadic amplification of the insulin-like growth factor 1 receptor gene in human breast tumors", Cancer Res. 52:1036-1039.
6. Bishop J.M. (1991), "Molecular themes in oncogenesis", Cell 64:235-248.
7. Blast R.C. Jr. (1993), "Perspectives on the future of cancer markers", Clin Chem. 31:2444-2451.
8. Brison O. (1993), "Gene amplification and tumor progression", Biochim. Biophys. Acta 1155:25-41.
9. Culver K.W. et al. (1994), "Gene therapy for cancer," Trends Genet. 10:174-178.
10. Henikoff et al. (1992), Proc. Natl. Acad. Sci. USA 89:10915-10919.
11. Kallioniemi A. et al. (1992), "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors", Science 258:818-821.
12. Kocher O. et al. (1995), "Identification of a novel gene, selectively up-regulated in human carcinomas, using the differential display technique", Clin. Cancer Res. 1:1209-1215.
13. Lippman M.E. (1993), "The development of biological therapies for breast cancer", Science 259:631-632.
14. MacLean G.D. et al. (1992), "The immune system, cancer antigens and immunotherapy", Contemp. Oncol. Aug/Sept.
15. McKenzie D. et al. (1994), "Using the RNA arbitrarily priumed polymerase chain reaction (RAP-PCR) to analyze gene expression in human breast cancer cells lines" [abstract], J. Cell. Biochem. 18D:248.
16. Muss H.B. et al. (1994), "c-*erb*B-2 expression and response to adjuvant therapy in women with node-positive early breast cancer", New Engl. J. Med. 330:1260-1266.
17. Morgan R.A. et al. (1993), "Human gene therapy," Annu. Rev. Biochem. 62:191-217.

18. Roth J.A. (1994), "Modulation of oncogene and tumor-suppressor gene expression: a novel strategy for cancer prevention and treatment", Ann. Surg. Oncol. 1:79-86.

19. Saint-Ruf C. et al. (1990), "Proto-oncogene amplification and homogeneously staining regions in human breast carcinomas", Genes Chromosomes Cancer 2:18-26.

20. Slamon D.J. et al. (1987), "Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene", Science 235:178-182.

21. Schwab M. et al. (1990), "Amplification of cellular oncogenes: a predictor of clinical outcome of human cancer", Genes Chromosomes Cancer 1:181-193.

22. Thompson C.T. et al. (1993), "Cytogenetic profiling using fluorescence in situ hybridization (FISH) and comparative genomic hybridization (CGH)", J. Cell. Biochem. 17G:139-143.

23. Unsigned (1994), "Synthetic vaccine stabilizes advanced cancer, prolongs survival", Oncol. News 3:1.

24. Watson M.A. et al. (1994), "Isolation of differentially expressed sequence tags from human breast cancer", Cancer Res. 54:4598-4602.

25. Watson M.A. et al. (1996), "Mammaglobulin, a mammary-specific member of the uteroglobulin gene family, is overexpressed in human breast cancer", Cancer Res. 56:860-865.

26. Zafrani B. et al. (1992), "Cytogenetic study of breast cancer", Hum Pathol 23:542-547.

Articles on Differential Display, RNA fingerprinting, and related techniques

1. Ayala M. et al. (1995), "New primer strategy improves precision of differential display", BioTechniques 18:842-850.

2. Bauer D. et al. (1993), "Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR), Nucl. Acids Res. 21:4272-4280.

3. Bertioli D.J. et al. (1995), "An analysis of differential display shows a strong bias towards high copy number mRNAs", Nucl. Acids Res. 23:4520-4523.

4. Chen Z. et al. (1995), "Differential expression of human tissue factor in normal mammary epithelial cells and in carcinomas", Molecular Med. 1:153-160.

5. Haag E. et al. (1994), "Effects of primer choice and source of *Taq* DNA polymerase on the bainding patterns of differential display RT-PCR", BioTechniques 17:226-228.

6. Hadman M. Et al. (1995), "Modifications to the differential display technique reduce background and increase sensitivity", Anal. Biochem. 226:383-386.

7. Ito T. et al. (1994), "Fluorescent differential display: arbitrarily primed RT-PCR fingerprinting on an automated DNA sequencer", FEBS Lett. 351:231-236.

8. Liang P. et al. (1992a), "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction", Science 257:967-971.

9. Liang P. et al. (1992b), "Differential display and cloning of messenger RNAs in human breast cancer versus mammary epithelial cells", Cancer Res. 52:6966-6968.

10. Liang P. et al. (1993), "Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization", Nucl. Acids Res. 21:3269-3275.

11. Liang P. et al. (1994), "Differential display using one-base anchored oligo-dT primers", Nucl. Acids Res. 22:5763-5764.

12. Liang P. et al. (1995a), "Recent advances in differential display", Curr. Opin. Immunol. 7:274-280.

13. Liang P. et al (1995b), "analysis of altered gene expression by differential display", Methods Enzymol. 254:304-321.

14. Linskens M.H.K. et al. (1995), "Cataloging altered gene expression in young and senescent cells using enhanced differential display", Nucl. Acids Res. 23:3244-3251.

15. Snager R. et al. (1993), "Identification by differential display of alpha-6 integrin as a candidate tumor suppressor gene", FASEB J. 7:964-970.

16. Sompayrac L. et al. (1995), "Overcoming limitations of the mRNA differential display technique", Nucl. Acids Res. 23:4738-4739.

17. Sun Y. et al. (1994), "Moelcular cloning of five messenger RNAs differentially expressed in preneoplastic or neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinases-3", Cancer Res. 54:1139-1144.

18. Sunday M.E. et al. (1995), "Differential display RT-PCR for identifying novel gene expression in the lung", Am. J. Physiol. 269:L273-L284.

19. Trentmann S.M. et al. (1995), "Alternatives to ^{35}S as a label for the differential display of eukaryotic messenger RNA", Science 267:1186-1187.

20. Welsh J. et. al. (1992), "Arbitrarily primed PCR fingerprinting of RNA", Nucl. Acids Res. 20:4965-4970.

21. Yeatman T.J. et al. (1995), "Identification of a differentially-expressed message associated with colon cancer liver metastasis using an improved method of differential display", Nucl. Acids Res. 23:4007-4008.

22. Yoshikawa T. et al. (1995), "Detection, simultaneous display and direct sequencing of multiple nuclear hormone receptor genes using bilaterally targeted RNA fingerprinting", Biochim. Biophys. Acta 1264:63-71.

Articles on Duplicated Chromosome Regions in Cancer

1. Bentz M. et al. (1994), "Fluorescent in situ hybridization in leukemias: 'the FISH are spawning!'", Leukemia 8:1447-1452.

2. Bentz M. et al. (1995a), "Comparative genomic hybridization in chronic B-cell leukemias shows a high incidence of chomosomal gains and losses", Blood 85:3610-3618.

3. Bentz M. et al. (1995b), "Comparative genomic hybridization in the investigation of myeloid leukemias", Genes Chrom. Cancer 12:193-200.

4. Bryndorf T. et al. (1995), "Comparative genomic hybridization in clinical cytogenetics", Am. J. Hu. Genetics 57:1211-1220.

5. Cher ML. et al. (1994), "Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer", Genes Chrom. Cancer 11:153-162.
6. Dutrillaux B. et al. (1990), "Characterization of chromosomal anomalies in human breast cancer", Cancer Genet Cytogenet 49:203-217.
7. Feuerstein BG. et al. (1995), "Molecular cytogenetic quantitation of gains and losses of genetic material from human gliomas", J. Neuro-Oncol. 24:47-55.
8. Forus A. et al. (1995a), "Comparative genomic hybridization analysis of human sarcomas: I. Occurrence of genomic imbalances and identification of a novel major amplicon at 1q21-q22 in soft tissue sarcomas", Genes Chrom. Cancer 14:8-14.
9. Forus A. et al. (1995b), "Comparative genomic hydridization analysis of human sarcomas: II. Identification of novel amplicons at 6p and 17p in osteosarcomas", Genes Chrom. Cancer 14:15-21.
10. Gordon KB. et al. (1994), "Comparative genomic hybridization in the detection of DNA copy number abnormalities in uveal melanoma", Cancer Res. 54:4764-4768.
11. Gray JW. et al. (1994), "Fluorescence in situ hybridization in cancer and radiation biology", Radiation Res. 137:275-289.
12. Houldsworth J. et al. (1994), "Comparative genomic hybridization: an overview", Am. J. Path. 145:1253-1260.
13. Isola JJ. et al. (1995), "Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer", Am. J. Path. 147: 905-911.
14. Iwabuchi H. et al. (1995), "Genetic analysis of benign, low-grade, and high-grade ovarian tumors", Cancer Res. 55:6172-6180.
15. Kallioniemi A. et al. (1994), "Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization", Proc. Natl. Acad. Sci. USA 91:2156-2160.
16. Kallioniemi A. et al. (1995), "Identification of gains and losses of DNA sequences in primary bladder cancer by comparative genomic hybridization", Genes Chrom. Cancer 12:213-219.
17. Kim DH. et al. (1995), "Chromosomal abnormalities in glioblastoma multiforme tumors and glioma cell lines detected by comparative genomic hybridization", Int. J. Cancer 60:812-819.
18. Levin NA. et al. (1994), "Identification of frequent novel genetic alterations in small cell lung carcinoma", Cancer Res. 5086-5091.
19. Levin NA. et al. (1995), "Identification of novel regions of altered DNA copy number in small cell lung tumors", Genes Chrom. Cancer 13:175-185.
20. Lisitsyn NA. et al. (1995), "Comparative genomic analysis of tumors: detection of DNA losses and amplification", Proc. Natl. Acad. Sci. USA 92:151-155.
21. Mohamed AN. et al. (1994), "Extrachromosomal gene amplification in acute myeloid leukemia; characterization by metaphase analysis, comparative genomic hybridization, and semi-quantitative PCR", Genes Chrom. Cancer 8:185-189.
22. Mohapatra G. et al. (1995), "Detection of multiple gains and losses of genetic material in ten glioma cell lines by comparative genomic hybridization" Genes Chrom. Cancer 13:86-93.

23. Muleris M. et al. (1994a), "Detection of DNA amplification in 17 primary breast carcinomas with homogeneously staining regions by a modified comparative genomic hybridization technique", Genes Chrom. Cancer 10:160-170.

24. Muleris M. et al. (1994b), "Oncogene amplification in human fliomas: a molecular cytogenetic analysis", Oncogene 9:2717-2722.

25. Nacheva E. et al. (1995), "Comparative genomic hybridization in acute myeloid leukemia. A comparison with G-banding and chromosome painting", Cancer Genetics Cytogenetics 82:9-16.

26. Ried T. et al. (1994), "Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization", Cancer Res. 54:1801-1806.

27. Ried T. et al. (1995), "Comparative genomic hybridization of formalin-fixed, paraffin-embedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas", Cancer Res. 55:5415-5423.

28. Schlegel J. et al. (1994), "Detection of amplified DNA sequences by comparative genomic in situ hybridization with human glioma tumor DNA as probe", Verhand. Deut. G. Path. 78:204-207.

29. Schlegel J. et al. (1995), "Comparative genomic in situ hybridization of colon carcinomas with replication error", Cancer Res. 55:6002-6005.

30. Schlegel J. et al. (1996), "Detection of complex genetic alterations in human glioblastoma multiforme using comparative genomic hybridization", J. Neuropà Mol. Exp. Neurol. 55:81-87.

31. Schrock E. et al. (1994), "Comparative genomic hybridization of human malignant gliomas reveals multiple amplification sites and nonrandom chromosomal gains and losses", Am. J. Path. 144:1203-1218.

32. Seruca R. et al. (1995), "Increasing levels of MYC and MET co-amplification during tumor progression of a case of gastric cancer", Cancer Genetics Cytogenetics 82:140-145.

33. Speicher MR. et al. (1994), "Chromosomal gains and losses in uveal melanomas detected by comparative genomic hybridization", Cancer Res. 54:3817-3823.

34. Speicher MR. et al. (1995), "Comparative genomic hybridization detects novel deletions and amplifications in head and neck squamous cell carcinomas", Cancer Res. 55:1010-1013.

35. Steilen-Gimbel H. et al. (1996), "A novel site of DNA amplification on chromosome 1p32-33 in a rhabdomyosarcoma revealed by comparative genomic hybridization", Hu. Genetics 97:87-90.

36. Suijkerbuijk RF. et al. (1994), "Comparative genomic hybridization as a tool to define two distinct chromosome 12-derived amplification units in well-differentiated liposarcomas", Genes Chrom Cancer 9:292-295.

37. Tanner MM. et al. (1994), "Increased copy number at 20q13 in breast cancer: defining the critical region and exclusion of candidate genes", Cancer Res. 54:4257-4260.

38. Tarkkanen M. et al. (1995), "Gains and losses of DNA sequences in osteosarcomas by comparative genomic hybridization", Cancer Res. 55:1334-1338.

39. Visakorpi T. et al. (1995a), "Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization", Cancer Res. 55:342-347.

40. Visakorpi T. et al. (1995b), "In vivo amplification of the androgen receptor gene and progression of human prostate cancer", Nature Genetics 9:401-406.

41. Voorter C. et al. (1995), "Detection of chromosomal imbalances in transitional cell carcinoma of the bladder by comparative genomic hybridization", Am. J. Path. 146:1341-1354.

42. Wiltshire RN. et al. (1995), "Direct visualization of the clonal progression of primary cutaneous melanoma: application of tissue microdissection and comparative genomic hybridization", Cancer Res. 55:3954-3957.

TABLE OF SEQUENCE LISTINGS:

SEQ. ID NO	Designation	Description	Type	
1	CH1-9a11-2	152 bp sequence (fragment)	dsDNA	Figure 22
2		translation	amino acid	Figure 22
3	CH8-2a13-1	163 bp sequence (fragment)	dsDNA	Figure 22
4		translation	amino acid	Figure 22
5	CH13-2a12-1	107 bp sequence (fragment)	dsDNA	Figure 22
6		translation	amino acid	Figure 22
7	CH14-2a16-1	114 bp sequence (fragment)	dsDNA	Figure 22
8		translation	amino acid	Figure 22
9 to 14	Primers		ssDNA	Table 2
15	CH1-9a11-2	3.5 kb nucleotide sequence	dsDNA	Figure 8
16		translation	amino acid	Figure 9
17		protein	amino acid	Figure 9
18	CH8-2a13-1	4.0 kb nucleotide sequence	dsDNA	Figure 11
19		translation	amino acid	Figure 12
20		protein	amino acid	Figure 12
21		4.1 kb sequence (predicted)	dsDNA	Figure 13
22		translation	amino acid	Figure 14
23	CH13-2a12-1	3.3 kb nucleotide equence	dsDNA	Figure 16
24		translation	amino acid	Figure 17
25		protein	amino acid	Figure 17
26	CH14-2a16-1	2.0 kb nucleotide sequence	dsDNA	Figure 19
27		translation	amino acid	Figure 20
28		protein	amino acid	Figure 20
29		0.6 kb nucleotide sequence	ssDNA	Figure 21
30		translation	amino acid	Figure 21

SEQ. ID NO	Designation	Description	Type	
31		0.3 kb nucleotide sequence	ssDNA	Figure 21
32		translation	amino acid	Figure 21
33	CH1-9a11-2	3.5 kb nucleotide sequence	dsDNA	Figure 23
34		translation	amino acid	Figure 24
35	CH14-2a16-1	2.0 kb nucleotide sequence	dsDNA	Figure 25
36		translation	amino acid	Figure 26
37		protein	amino acid	Figure 26
38 & 39	Motif	Zinc-finger binding domain	dsDNA	text
40-43	Primers		ssDNA	text
44 & up	Primers		ssDNA	Figures 7, 10, 15, 18

SEQ ID NO:9:
TTTTTTTTTT TCC 13

SEQ ID NO:10:
TTTTTTTTTT TAC 13

SEQ ID NO:11:
CAATCGCCGT 10

SEQ ID NO:12:
TCGGCGATAG 10

SEQ ID NO:13:
CAGCACCCAC 10

SEQ ID NO:14:
AGCCAGCGAA 10

CLAIMS

What is claimed as the invention is:

1. An isolated polynucleotide comprising a linear sequence of at least 10 nucleotides identical to a linear sequence contained in a polynucleotide selected from the group consisting of CH8-2a13-1, CH13-2a12-1, CH14-2a16-1, and CH1-9a11-2.

2. An isolated polynucleotide comprising a linear sequence of at least 40 consecutive nucleotides at least 90% identical to a linear sequence contained in a sequence selected from the group consisting of SEQ. ID NO:15, SEQ. ID NO:18, SEQ. ID NO:21, SEQ. ID NO:23, SEQ. ID NO:26, SEQ. ID NO:29, SEQ. ID NO:31,. SEQ. ID NO:33, and SEQ. ID NO:35; but not in any of SEQ. ID NOS: 1, 3, 5, and 7.

3. The isolated polynucleotide of claim 2, comprising a linear sequence of at least 100 consecutive nucleotides at least 90% identical to a sequence contained in the selected sequence.

4. The isolated polynucleotide of claim 2, comprising a linear sequence of at least 40 consecutive nucleotides at least 95% identical to a sequence contained in the selected sequence.

5. An isolated polynucleotide comprising a linear sequence of at least 40 consecutive nucleotides that hybridizes with a DNA having a sequence selected from the group consisting of SEQ. ID NO:15, SEQ. ID NO:18, SEQ. ID NO:21, SEQ. ID NO:23, SEQ. ID NO:26, SEQ. ID NO:29, SEQ. ID NO:31,. SEQ. ID NO:33, and SEQ. ID NO:35; under conditions where it does not hybridize with SEQ. ID NOS: 1, 3, 5, 7, or any other DNA from a human cell.

6. The isolated polynucleotide of claim 5, wherein the linear sequence is at least 100 consecutive nucleotides

7. An isolated polynucleotide comprising a sequence of at least 40 consecutive nucleotides that hybridizes with an RNA having a sequence selected from the group consisting of SEQ. ID NO:15, SEQ. ID NO:18, SEQ. ID NO:21, SEQ. ID NO:23, SEQ. ID NO:26, SEQ. ID NO:29, SEQ. ID NO:31,. SEQ. ID NO:33, and SEQ. ID NO:35; under conditions where it does not hybridize with SEQ. ID NOS: 1, 3, 5, 7, or any other RNA from a human cell.

8. The isolated polynucleotide of claim 7, wherein the linear sequence is at least 100 consecutive nucleotides

9. The isolated polynucleotide of any of claims 2-8, wherein said linear sequence is contained in a duplicated gene or overabundant RNA in cancerous cells.

10. The isolated polynucleotide of any of claims 2-8, which is a CH13-2a12-1 polynucleotide, and is contained in an encoding region for a protein or RNA molecule that controls cell proliferation.

11. The isolated polynucleotide of any of claims 2-8, which is a CH14-2a16-1 polynucleotide, and is contained in an encoding region for a protein with DNA or RNA binding activity.

12. The isolated polynucleotide of any of claims 2-8, present in a recombinant plasmid deposited under ATCC Accession No. 98074

13. The isolated polynucleotide of any of claims 2-8, present in a recombinant phage deposited under ATCC Accession No. 97595.

14. The isolated polynucleotide of any of claims 2-8, present in the λBCBT474 cDNA library deposited under ATCC Accession No. 97594.

15. An isolated polynucleotide comprising a linear sequence of polynucleotides essentially identical to a sequence selected from the group consisting of SEQ. ID NO:15, SEQ. ID NO: 18, SEQ. ID NO:21, SEQ. ID NO:23, SEQ. ID NO:26, SEQ. ID NO:29, SEQ. ID NO:31, SEQ. ID NO:33, and SEQ. ID NO:35.

16. An isolated polypeptide comprising a linear sequence of at least 5 amino acid residues identical to a sequence encoded by a polynucleotide selected from the group consisting of CH1-9a11-2, CH8-2a13-1, CH13-2a12-1, and CH14-2a16-1.

17. An isolated polypeptide comprising a linear sequence of at least 5 consecutive amino acids identical to a linear sequence contained in a sequence selected from the group consisting of SEQ. ID NO:17, SEQ. ID NO:20, SEQ. ID NO:22, SEQ. ID NO:24, SEQ. ID NO:28, SEQ. ID NO:30, SEQ. ID NO:32, SEQ. ID NO:34, and SEQ. ID NO:37; but not in any of SEQ. ID NOS: 2, 4, 6, and 8.

18. The isolated polypeptide of claim 17, comprising a linear sequence of at least 15 consecutive amino acids at least 90% identical to a linear sequence contained in the selected sequence.

19. The isolated polypeptide of claim 17 or 18, wherein said linear sequence is encoded in a duplicated gene or overabundant RNA in cancerous cells.

20. The isolated polypeptide of claim 17 or 18, which is overexpressed in cancerous cells.

21. The isolated polypeptide of claim 17 or 18, wherein the polynucleotide selected from said group is a CH1-9a11-2 polynucleotide, and the polypeptide is a transmembrane polypeptide.

22. An isolated polypeptide comprising a linear sequence of amino acids essentially identical to a sequence selected from the group consisting of SEQ. ID NO:17, SEQ. ID NO:20, SEQ. ID NO:22, SEQ. ID NO:24, SEQ. ID NO:28, SEQ. ID NO:30, SEQ. ID NO:32, SEQ. ID NO:34, and SEQ. ID NO:37; but not in any of SEQ. ID NOS: 2, 4, 6, and 8.

23. An isolated polynucleotide comprising an encoding sequence for the polypeptide of any of claims 17 to 22.

24. A monoclonal or isolated polyclonal antibody specific for the polypeptide of claim 22.

25. A method of detecting gene duplication in cancerous cells, comprising the steps of:
   a) reacting DNA contained in a clinical sample with a reagent comprising the polynucleotide of claims 2-8, said clinical sample having been obtained from an individual suspected of having cancerous cells; and
   b) comparing the amount of any complexes formed between the reagent and the DNA in the clinical sample with the amount of any complexes formed between the reagent and DNA in a control sample.

26. A method of detecting overabundance of RNA in cancerous cells, comprising the steps of:
   a) reacting RNA contained in a clinical sample with a reagent comprising the polynucleotide of claim 2-8, said clinical sample having been obtained from an individual suspected of having cancerous cells; and
   b) comparing the amount of any complexes formed between the reagent and the RNA in the clinical sample with the amount of any complexes formed between the reagent and RNA in a control sample.

27. A method of determining gene duplication or overabundance of RNA in cancerous cells, comprising the steps of:

a) amplifying DNA or RNA in a clinical sample with a primer comprising the polynucleotide of claim 2-8 to yield an amplified polynucleotide, said clinical sample having been obtained from an individual suspected of having cancerous cells; and

b) comparing the amount of polynucleotide amplified from the DNA or RNA with the amount of polynucleotide amplified from DNA or RNA from a control sample.

28. A method of screening for cancer associated with a gene duplication in an individual, comprising the steps of:

a) determining gene duplication in cells from the individual according to the method of claim 25; and

b) correlating any gene duplication determined in step a) with an increased risk for the cancer.

29. A method of screening for cancer associated with overexpression of RNA in an individual, comprising the steps of:

a) determining overexpression of RNA in cells from the individual according to the method of claim 26; and

b) correlating any RNA overexpression determined in step a) with an increased risk for the cancer.

30. A method of screening for cancer associated with a gene duplication or overexpression of RNA in an individual, comprising the steps of:

a) determining gene duplication or overexpression of RNA in cells from the individual according to the method of claim 27; and

b) correlating any gene duplication or overexpression of RNA determined in step a) with an increased risk for the cancer.

31. The method of any of claims 28-30, which is a screening method for breast cancer.

32. A diagnostic kit for detecting gene duplication or RNA overabundance in cells contained in an individual as manifest in a clinical sample, comprising a reagent and a buffer in suitable packaging, wherein the reagent comprises the polynucleotide of any of claims 2-8.

33 A method for detecting altered protein expression in cancerous cells, comprising the steps of:

a) reacting a polypeptide contained in a clinical sample with a reagent comprising the antibody of claim 24, said clinical sample having been obtained from an individual suspected of having cancerous cells; and

b) comparing the amount of any complexes formed between the reagent and the polypeptide in the clinical sample with the amount of any complexes formed between the reagent and a polypeptide in a control sample.

34. A diagnostic kit for detecting a polypeptide present in a clinical sample, comprising a reagent and a buffer in suitable packaging, wherein the reagent comprises the antibody of claim 24.

35. A host cell genetically altered by the polynucleotide of any of claims 2 to 8 or claim 23.

36. A method of screening a pharmaceutical candidate, comprising the steps of:

a) separating progeny of the cell of claim 35 into a first group and a second group;

b) treating the first group of cells with the pharmaceutical candidate;

c) not treating the second group of cells with the pharmaceutical candidate; and

d) comparing the phenotype of the treated cells with that of the untreated cells.

37. A pharmaceutical preparation for use in cancer therapy, comprising the polynucleotide of claim 2 to 8 or claim 23, said preparation being capable of reducing the pathology of cancerous cells.

38. A method for treating an individual bearing cancerous cells, comprising administering the pharmaceutical preparation of claim 37.

39. A pharmaceutical preparation for use in cancer therapy, comprising the antibody of claim 24, said preparation being capable of reducing the pathology of cancerous cells.

40. A method for treating an individual bearing cancerous cells, comprising administering the pharmaceutical preparation of claim 39.

41. A pharmaceutical preparation comprising the polypeptide of claim 17 or 18 in an immunogenic form, and a pharmaceutically compatible excipient.

42. A method for treatment of cancer, comprising administration of the pharmaceutical preparation of claim 41.

43. A method for obtaining cDNA corresponding to a gene that is duplicated or overexpressed in cancer, comprising the steps of:

a) supplying an RNA preparation from control cells;

b) supplying RNA preparations from at least two different cancer cells;

c) displaying cDNA corresponding to the RNA preparations of step a) and step b) such that different cDNA corresponding to different RNA in each preparation are displayed separately;

d) selecting cDNA corresponding to RNA that is present in greater abundance in the cancer cells of step b) relative to the control cells of step a);

e) supplying a digested DNA preparation from control cells;

f) supplying digested DNA preparations from at least two different cancer cells;

g) hybridizing the cDNA of step d) with the digested DNA preparations of step e) and step f); and

h) further selecting cDNA from the cDNA of step d) corresponding to a gene that is duplicated in the cancer cells of step f) relative to the control cells of step e).

44. The method of claim 43, wherein the two different cancer cells used to supply RNA in step b) share a duplicated gene in the same region of a chromosome.

45. The method of claim 43, wherein RNA preparations from at least three different cancer cells are supplied in step b).

46. The method of claim 43, wherein the three different cancer cells used to supply RNA in step b) share a duplicated gene in the same region of a chromosome.

47. The method of claim 43, wherein the control cells of step a) are uncultured.

48. The method of claim 43, further comprising supplying a digested mitochondrial DNA preparation; hybridizing the cDNA of step h) with the digested mitochondrial DNA preparation; and further selecting cDNA from the cDNA of step h) corresponding to genes that do not hybridize with the digested mitochondrial DNA preparation.

49. The method of claim 43, further comprising the steps of:

i) supplying an RNA preparation from control cells;

j) supplying RNA preparations from at least two different cancer cells;

k) hybridizing the cDNA of step h) with the RNA preparations of step i) and step j); and

l) further selecting cDNA from the cDNA of step h) corresponding to RNA that is present in greater abundance in the cancer cells of step j) relative to the control cells of step i).

50. The method of claim 49, wherein the gene to which the cDNA corresponds is not duplicated in at least one of the cancer cells used to supply the RNA in step j) relative to the control cells of step e).

51. The method of claim 43, wherein the two different cancer cells used to supply the RNA preparations in step b) are breast cancer cells.

52. The method of claim 43, wherein the two different cancer cells used to supply the RNA preparations in step b) are from a common type of cancer, wherein the type of cancer is selected from the group consisting of lung cancer, glioblastoma, pancreatic cancer, colon cancer, prostate cancer, hepatoma, and myeloma.

53. The method of claim 43, wherein the two different cancer cells used to supply the digested DNA preparations in step f) are breast cancer cells.

54. The method of claim 43, wherein the two different cancer cells the digested DNA preparations in step f) are from a common type of cancer, wherein the type of cancer is selected from the group consisting of lung cancer, glioblastoma, pancreatic cancer, colon cancer, prostate cancer, hepatoma, and myeloma.

55. A method for obtaining cDNA corresponding to a gene that is deleted or underexpressed in cancer, comprising the steps of:

a) supplying an RNA preparation from control cells;

b) supplying RNA preparations from at least two different cancer cells that share a deleted gene in the same region of a chromosome;

c) displaying cDNA corresponding to the RNA preparations of step a) and step b) such that different cDNA corresponding to different RNA in each preparation are displayed separately; and

d) selecting cDNA corresponding to RNA that is present in lower abundance in the cancer cells of step b) relative to the control cells of step a).

56. The method of claim 55, further comprising the steps of:

e) supplying a digested DNA preparation from control cells;

f) supplying digested DNA preparations from at least two different cancer cells;

g) hybridizing the cDNA of step d) with the digested DNA preparations of step e) and step f); and

h) further selecting cDNA from the cDNA of step d) corresponding to a gene that is deleted in the cancer cells of step f) relative to the control cells of step e).

57. A method for characterizing a gene that is duplicated or has altered expression in cancer, comprising obtaining cDNA corresponding to the gene according to the method of any of claims 43-56, and then sequencing the cDNA.

58. A method of screening a candidate drug for cancer treatment, comprising obtaining cDNA corresponding to a gene that is duplicated or has altered expression in cancer according to the method of any of claims 43-56, and comparing the effect of the candidate drug on a cell genetically altered with the cDNA with the effect on a cell not genetically altered with the cDNA.

Figure 1

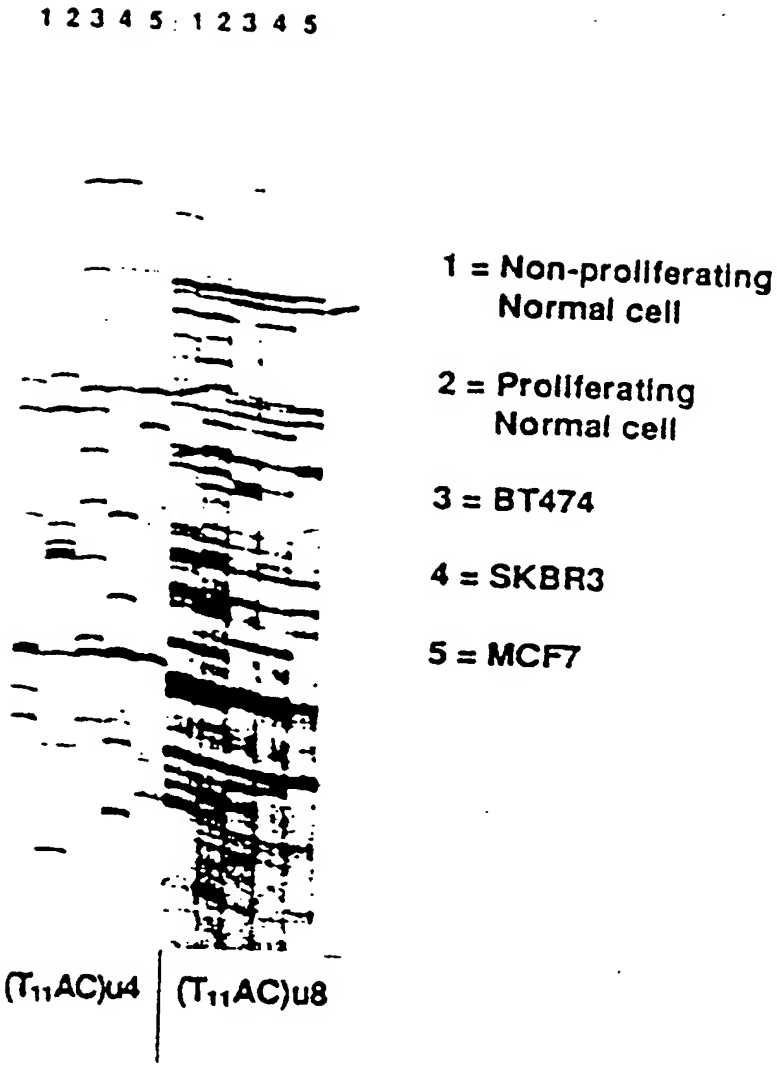
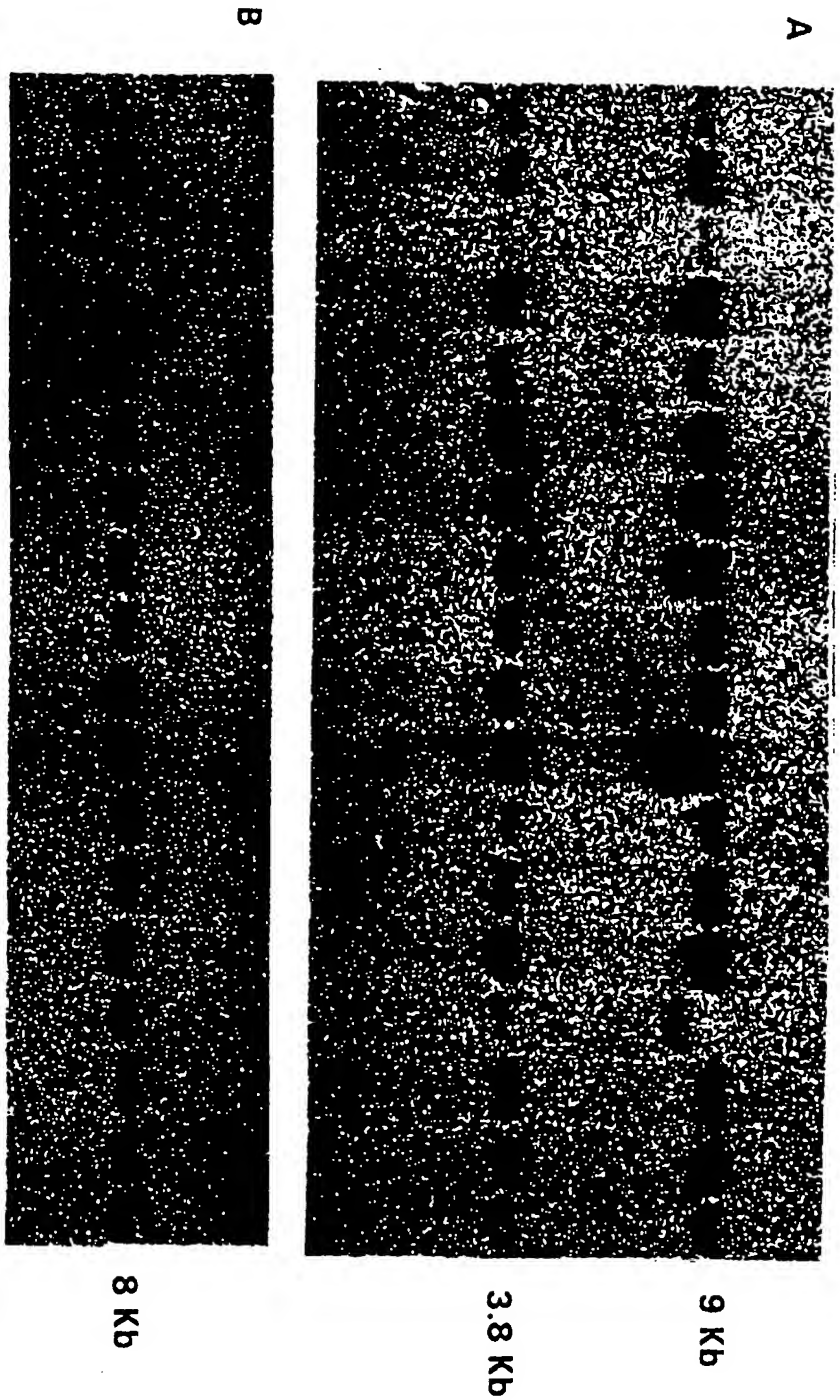
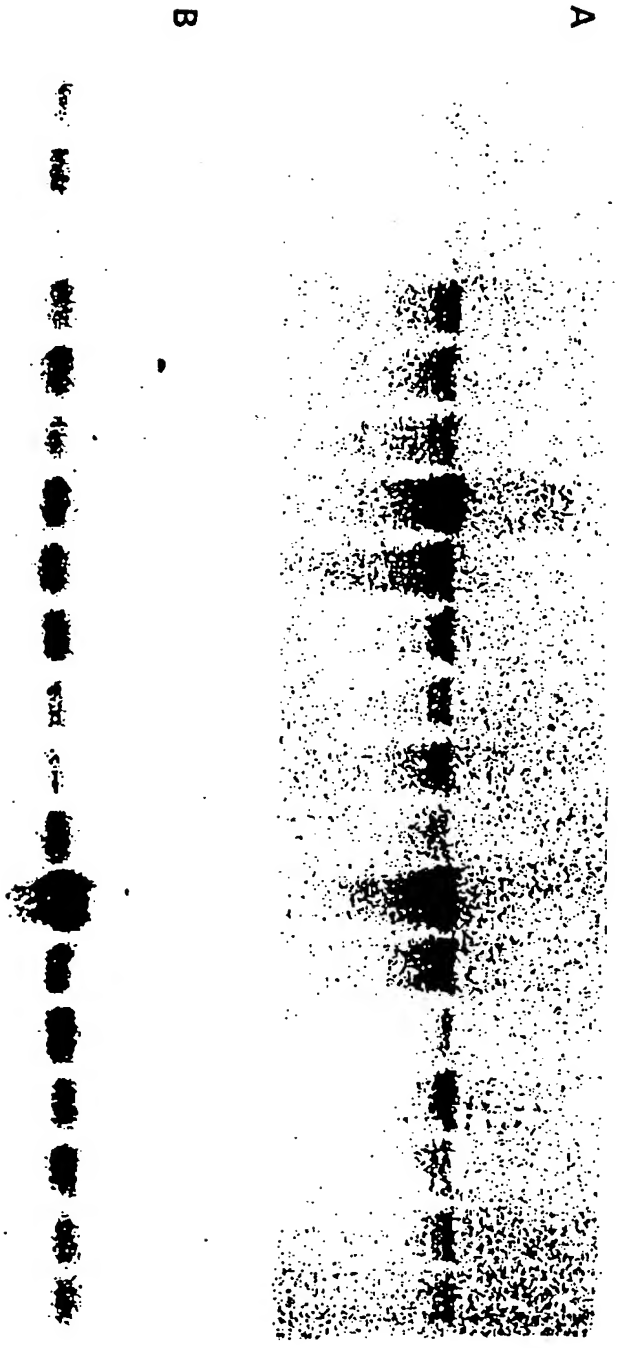

1 2 3 4 5 1 2 3 4 5
1 = Non-proliferating Normal cell
2 = Proliferating Normal cell
3 = BT474
4 = SKBR3
5 = MCF7
(T11AC)u4
(T11AC)u8

Figure 2



1: P11
2: P12
3: G94 SKF
4: BT 484
5: MCF 7
6: MDA-MB-157
7: SKBR 3
8: ZR-75-30
9: CAMA-1
10: DU 4475
11: 600 MPE
12: MDA-MB-453
13: MDA-MB-231
14: T47D
15: UACC 812
16: BT 468
17: MDA-MB-134
18: MDA-MB-435

Figure 3



1: 415 ER
2: 640 ERNP
3: 640 ERP
4: BT 474
5: CAMA 1
6: SKBR 3
7: 600 MPE
8: MDA-MB-361
9: MDA-MB-231
10: MDA-MB-134
11: MDA-MB-468
12: BT 20
13: T47D
14: MDA-MB-157
15: UACC 812
16: MDA-MB-435
17: MCF 7
18: MDA-MB-453
19: DU 4475

Figure 4

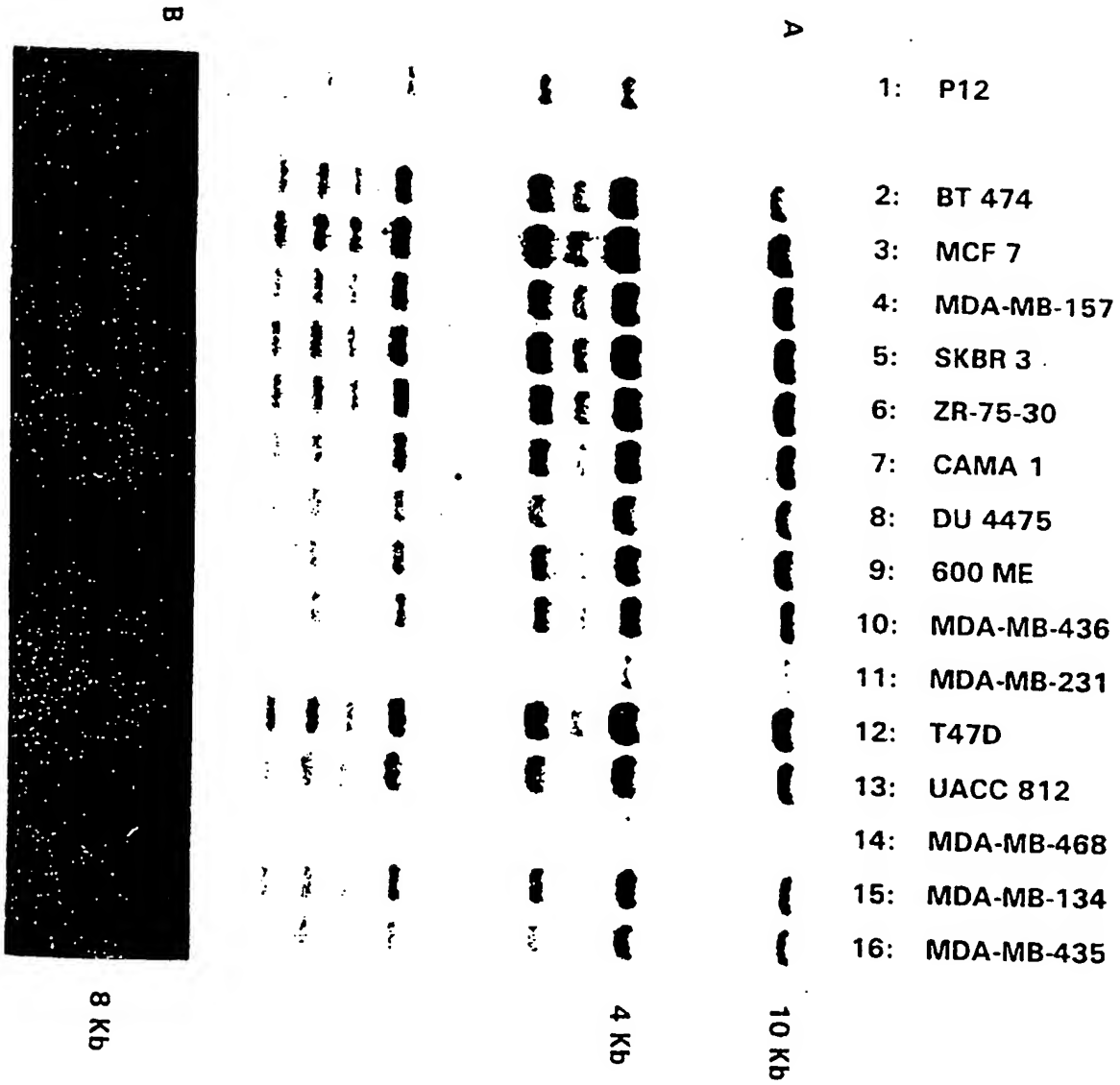
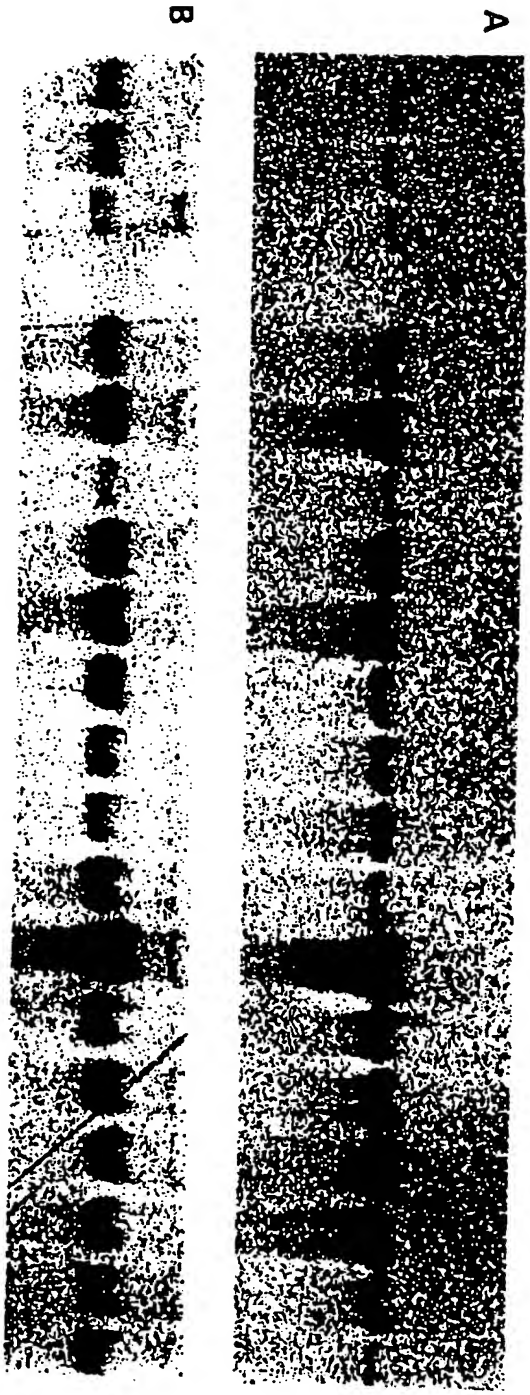

B
A
8 Kb
4 Kb
10 Kb
1: P12
2: BT 474
3: MCF 7
4: MDA-MB-157
5: SKBR 3
6: ZR-75-30
7: CAMA 1
8: DU 4475
9: 600 ME
10: MDA-MB-436
11: MDA-MB-231
12: T47D
13: UACC 812
14: MDA-MB-468
15: MDA-MB-134
16: MDA-MB-435

Figure 5

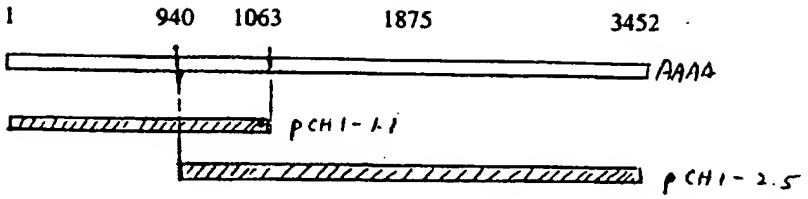


1: 415 ER
2: 640 ERNP
3: 640 ERP

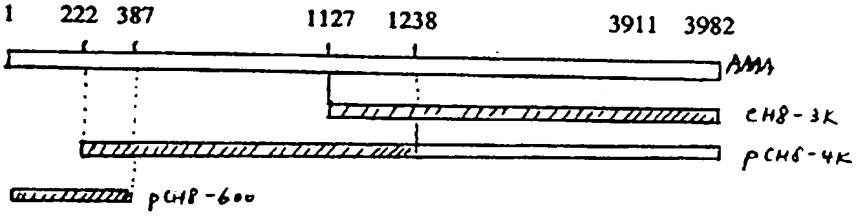
4: BT 474
5: CAMA 1
6: SKBR 3
7: 600 MPE
8: MDA-MB-361
9: MDA-MB-231
10: MDA-MB-134
11: MDA-MB-468
12: BT 20
13: T47D
14: MDA-MB-157
15: UACC 812
16: MDA-MB-435
17: MCF 7
18: MDA-MB-453
19: DU 4475

Figure 6

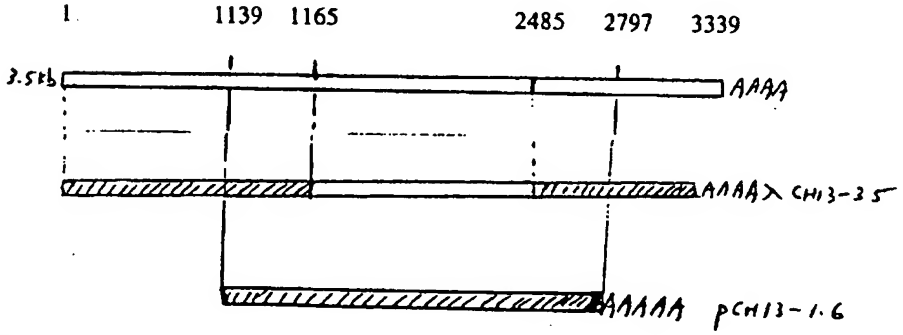
CH1-9a11-2



CH8-2a13-1



CH13-2a12-1



CH14-2a16-1

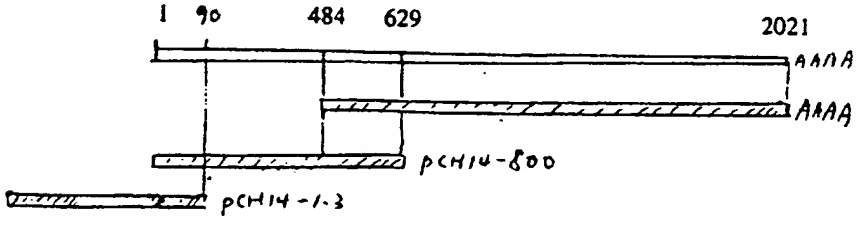


Figure 7

+ strand (sense)	sequence (5'-->3')	
	1st base	
1. pch1-t7-1f	1123	CGG GAG GTT TCA GAT CGA C
2. pch1-t7-2f	1437	GCG CTG CAA GTA CAA AAT TG
3. pch1-t7-3f	1729	TCT AAA GTC CAA GAC CAA GG
4. pch1-t7-4f	1987	CAG AAA TTA TGG TTT CTA CC
5. pch1-t7-5f	2266	CaG GAA GAG GAG GGA TAA C
6. pch1-sp6-3fb	2684	(T) AAA CAT ACA CAA TAA ACA C
7. pch1-sp6-2rb	2966	TTG GCA GCG ACT GTA TTT G
8. pch1-sp6-1rb	3283	CCT GAT TTT ATA GAA GCC CC
- strand (antisense)		
9. pch1-sp6-1f	3302	GGG GCT TCT ATA AAA TCA GG
10. pch1-sp6-2f	2987	ATT CAA ATA CAG TTG CTG C
11. pch1-sp6-3f	2705	TTA GTG TTT ATT GTG TAT G
12, pch1-sp6-4f	2458	AGT GTT CAT TTC CAG TGA G
13, pch1-sp6-5f	2066	CTT TGT TCT TGG ACT TTA G
14. pch1-t7-3fb	1748	CCT TGG TCT TGG ACT TTA G
15. pch1-t7-2rb	1445	AAT TTT GTA CTT GCA GCG C
16. pch1-t7-1rb	1141	GTC GAT CTG AAA CCT CCC G
17. CH1a	1063	GTG CCT GTA GCA ACT GGA TGG C
18. CH1b	1079	GTC ATG TTG GTC AGC TGT GCC

Figure 8(A)

```
   1  GAATACATAT ATAAATGGTG TTCAGTTAGA GTTGCTCTTT ATCGGCAGCG
  51  CAGCCGAACT GCTTTGAGTA AAGGAAAAGA TTATCTTGTG TTAGCTCAAC
 101  CACCCTTACT ACTTCCTGCG GAATCAGTAG ATGTTTCAGT ATTGCAACCT
 151  CTGAGTGGAG AATTGGAAAA TACGAATATA GAAAGGGAAG CTGAAACTGT
 201  TGTTCTGGGT GATTTAAGTA GTAGTATGCA CCAGGATGAC TTGGTGAATC
 251  ACACTGTAGA TGCAGTTGAA CTTGAACCAA GCCATTCTCA AACTCTTTCT
 301  CAGTCTCTTC TTTTAGATAT TACCCCAGAA ATCAATCCCT TGCCTAAAAT
 351  AGAAGTATCT GAGTCTGTTG AATATGAGGC AGGACATATA CCATCACCAG
 401  TGATTCCCCA AGAGAGTTCT GTTGAGATCG ATAATGAAAC AGAACAAAAG
 451  TCTGAGAGCT TTAGTTCTAT AGAGAAACCA TCTATTACCT ATGAAACAAA
 501  TAAAGTTAAT GAGTTAATGG ATAATATTAT AAAAGAAGAT ATGAACTCCA
 551  TGCAAATTTT CACAAAGCTG TCTGAAACAA TAGTGCCACC AATAAATACA
 601  GCCACTGTAC CCGACAATGA AGATGGGGAA GCCAAAATGA ATATAGCTGA
 651  CACAGCAAAG CAAACTTTGA TTTCTGTTGT GGATTCTTCT TCATTACCTG
 701  AAGTAAAAGA AGAAGAACAG TCTCCAGAAG ATGCCCTTTT GAGAGGGTTA
 751  CAGAGGACAG CTACAGATTT TTATGCTGAA TTGCAAAATT CTACAGATCT
 801  AGGATATGCT AATGGAAATC TTGTACATGG ATCAAACCAA AAGGAGTCAG
 851  TATTTATGAG ACTTAATAAT CGTATTAAAG CCTTAGAAGT TAACATGTCT
 901  CTCAGTGGTC GCTATCTGGA GGAGCTTAGC CAAAGGTACC GAAAACAAAT
 951  GGAAGAAATG CAAAAGGCTT TCAACAAAAC AATCGTGAAA CTTCAGAATA
1001  CTTCAAGAAT AGCAGAGGAG CAGGATCAGC GGCAAACTGA AGCCATCCAG
1051  TTGCTACAGG CACAGCTGAC CAACATGACA CAGCTTGTTT CAAATTTATC
1101  AGCAACAGTA GCAGAATTGA AACGGGAGGT TTCAGATCGA CAAAGCTATC
1151  TTGTCATATC TTTGGTTCTT TGTGTTGTCT TGGGACTGAT GCTTTGTATG
1201  CAGCGTTGTC GAAATACTTC TCAATTTGAT GGAGATTATA TTTCAAAACT
1251  TCCTAAAAGT AATCAGTATC CAAGCCCTAA AAGGTGTTTC TCTTCCTATG
1301  ATGATATGAA TTTGAAAAGA AGAACTTCAT TCCCACTCAT GAGATCCAAG
1351  TCTCTACAGT TAACTGGCAA AGAAGTAGAC CCAAATGATT TGTACATTGT
1401  AGAACCCCTC AAGTTTTCTC CAGAAAAGAA GAAGAAGCGC TGCAAGTACA
1451  AAATTGAAAA AATTGAGACC ATAAAGCCTG AAGAACCATT GCACCCCATA
1501  GCCAATGGCG ACATAAAAGG AAGAAAGCCC TTTACGAACC AGAGAGATTT
1551  TTCTAATATG GGAGAAGTTT ATCACTCTTC TTATAAAGGT CCTCCATCTG
1601  AAGGAAGCTC AGAAACTTCA TCACAGTCAG AAGAGTCCTA TTTTTGTGGC
1651  ATTTCAGCTT GCACAAGTCT GTGCAATGGA CAGTCTCAAA AGACAAAAAC
1701  TGAGAAGAGG GCTTTAAAAC GAAGACGATC TAAAGTCCAA GACCAAGGAA
1751  AATTGATAAA AACTCTAATA CAGACTAAGT CGGGATCATT GCCGAGCCTG
1801  CATGACATAA TCAAAGGAAA CAAAGAGATC ACCGTGGGAA CATTTGGTGT
1851  TACAGCAGTC TCGGGACATA TCTAAAATTA ATTGAACTTT TCATACAGAA
1901  GACTTTTTTG TTGTTGTTCT TTGAAGAACA GTCTGTAGTA TTTGAAGGGT
1951  TTGGGGGAGG GAGAAAATAT TAATGGGAAA GGCATTCAGA AATTATGGTT
2001  TCTACCTTTT TAAAAAGTAG ATGGGATTGT GCTCAATCTT GGTTAATGAG
2051  CTACAGTTTT ACAAAGCTGA TCACTTCCTA TAAGGACAAT GGTAGACATT
2101  TTATAAAGAT GTTTTTTCAC AAGATTAATT ACTGGGACAA AAGTAATTTG
2151  GAAGCCCAGT TCCTTAGGTG GGATAGGAAT GAAAGCCTAA ACCTCTTCCT
2201  TTAGCTTTGT TCCTATTTCT TGCACCTTCC CATATTTATG TGCCTTTTGT
2251  CTATTTATAA TGCCACTGGA AGAGGAGGGA TAACTTTTTC TGTTATTTGA
2301  TTTCTTTTAT AACTTTGTTA GGTTTTTGAA GCTGCAAACA CTACAATGCT
2351  TTGAGGGGGT CTGTGCCTGA AGCTCAGGAG TGTGGATCAG ACAGTCTAAA
2401  GATCCTAAAA ACTTGCCAAC TGGATCTTTG TTTAGCAAAC TCACTGGAAA
2451  TGAACACTTA ATGGAATTTT TAAGTCTGTT CTGTTAGGTA GATGGTGATG
2501  CTCTTGTTAT TTTCACTTAT TCAGGCTGGA TTACTTCTTA CTTAGTTACT
2551  AACTCAATGA GGAAAAAATC CCTACAGGAT CTTTTTTTGC AAACAACTGA
```

Figure 8(B)

```
2601    TATATGCAGA CAAATTTTTG ACAAATTCAC CTTTTAAACA CGACGTTAAC
2651    CGATTTGTGA AGGTTTTCTT TAGCTTACAT TTTAAACATA CACAATAAAC
2701    ACTAATCCTC CAAACTTTCA CTGTTTTTAT TAGTATGAAT ATAAAATTTG
2751    AAGGTTTGGC CAATTAGTAC AAGTCTCATG ATATAATCAC AGCCTGCATA
2801    CATATGCACA GATCCAGTTA GTGAGTTTGT CAAGCTTAAT CTAATTGGTT
2851    AAGTCTAAAG AGATTATTAT TCCTTGATGT TTGCTTTGTA TTGGCTACAA
2901    ATGTGCAGAG GTAATACATA TGTGATGTCG ATGTCTCTGT CTTTTTTTTT
2951    GTCTTTAAAA AATAATTGGC AGCAACTGTA TTTGAATAAA ATGATTTCTT
3001    AGTATGATTG TACAGTAATG AATGAAAGTG GAACATGTTT CTTTTTGAAA
3051    GGGAGAGAAT TGACCATTTA TTGTTGTGAT GTTTAAGTTA TAACTTATTG
3101    AGCACTTTTA GTAGTGATAA CTGTTTTTAA ACTTGCCTAA TACCTTTCTT
3151    GGGTATTGTT TGTAATGTGA CTTATTTAAC GCCTTCTTTG TTTGTTTAAG
3201    TTGCTGCTTT AGGTTAACAG CGTGTTTTAG AAGATTTAAA TTTCTTTCCT
3251    GTCTGCACAA TTAGCTATTC AGAGCAAGAG GGCCTGATTT TATAGAAGCC
3301    CCTTGAAAAG AGGTCCAGAT GAGAGCAGAG ATACAGTGAG AAATTATGTG
3351    ATCTGTGTGT TGTGGGAAGA GAATTTTCAA TATGTAACTA CGGAGCTGTA
3401    GTGCCATTAG AAACTGTGAA TTTCCAAATA AATCTGAACA CTTGTCTTTA
3451    TT
```

Figure 9

```
   1  EYIYKWCSVR VALYRQRSRT ALSKGKDYLV LAQPPLLLPA ESVDVSVLQP
  51  LSGELENTNI EREAETVVLG DLSSSMHQDD LVNHTVDAVE LEPSHSQTLS
 101  QSLLLDITPE INPLPKIEVS ESVEYEAGHI PSPVIPQESS VEIDNETEQK
 151  SESFSSIEKP SITYETNKVN ELMDNIIKED MNSMQIFTKL SETIVPPINT
 201  ATVPDNEDGE AKMNIADTAK QTLISVVDSS SLPEVKEEEQ SPEDALLRGL
 251  QRTATDFYAE LQNSTDLGYA NGNLVHGSNQ KESVFMRLNN RIKALEVNMS
 301  LSGRYLEELS QRYRKQMEEM QKAFNKTIVK LQNTSRIAEE QDQRQTEAIQ
 351  LLQAQLTNMT QLVSNLSATV AELKREVSDR QSYLVISLVL CVVLGLMLCM
 401  QRCRNTSQFD GDYISKLPKS NQYPSPKRCF SSYDDMNLKR RTSFPLMRSK
 451  SLQLTGKEVD PNDLYIVEPL KFSPEKKKKR CKYKIEKIET IKPEEPLHPI
 501  ANGDIKGRKP FTNQRDFSNM GEVYHSSYKG PPSEGSSETS SQSEESYFCG
 551  ISACTSLCNG QSQKTKTEKR ALKRRRSKVQ DQGKLIKTLI QTKSGSLPSL
 601  HDIIKGNKEI TVGTFGVTAV SGHI•N•LNF SYRRLFCCCS LKNSL•YLKG
 651  LGEGENINGK GIQKLWFLPF •KVDGIVLNL G••ATVLQS• SLPIRTMVDI
 701  L•RCFFTRLI TGTKVIWKPS SLGGIGMKA• TSSFSFVPIS CTFPYLCAFC
 751  LFIMPLEEEG •LFLLFDFFY NFVRFLKLQT LQCFEGVCA• SSGVWIRQSK
 801  DPKNLPTGSL FSKLTGNEHL MEFLSLFC•V DGDALVIFTY SGWITSYLVT
 851  NSMRKKSLQD LFLQTTDICR QIFDKFTF•T RR•PICEGFL •LTF•TYTIN
 901  TNPPNFHCFY •YEYKI•RFG QLVQVS•YNH SLHTYAQIQL VSLSSLI•LV
 951  KSKEIIIP•C LLCIGYKCAE VIHM•CRCLC LFFCL•KIIG SNCI•IK•FL
1001  SMIVQ••MKV EHVSF•KGEN •PFIVVMFKL •LIEHF•••• LFLNLPNTFL
1051  GYCL•CDLFN AFFVCLSCCF RLTACFRRFK FLSCLHN•LF RARGPDFIEA
1101  P•KEVQMRAE IQ•EIM•SVC CGKRIFNM•L RSCSAIRNCE FPNKSEHLSL
```

```
   1  EYIYKWCSVR VALYRQRSRT ALSKGKDYLV LAQPPLLLPA ESVDVSVLQP
  51  LSGELENTNI EREAETVVLG DLSSSMHQDD LVNHTVDAVE LEPSHSQTLS
 101  QSLLLDITPE INPLPKIEVS ESVEYEAGHI PSPVIPQESS VEIDNETEQK
 151  SESFSSIEKP SITYETNKVN ELMDNIIKED MNSMQIFTKL SETIVPPINT
 201  ATVPDNEDGE AKMNIADTAK QTLISVVDSS SLPEVKEEEQ SPEDALLRGL
 251  QRTATDFYAE LQNSTDLGYA NGNLVHGSNQ KESVFMRLNN RIKALEVNMS
 301  LSGRYLEELS QRYRKQMEEM QKAFNKTIVK LQNTSRIAEE QDQRQTEAIQ
 351  LLQAQLTNMT QLVSNLSATV AELKREVSDR QSYLVISLVL CVVLGLMLCM
 401  QRCRNTSQFD GDYISKLPKS NQYPSPKRCF SSYDDMNLKR RTSFPLMRSK
 451  SLQLTGKEVD PNDLYIVEPL KFSPEKKKKR CKYKIEKIET IKPEEPLHPI
 501  ANGDIKGRKP FTNQRDFSNM GEVYHSSYKG PPSEGSSETS SQSEESYFCG
 551  ISACTSLCNG QSQKTKTEKR ALKRRRSKVQ DQGKLIKTLI QTKSGSLPSL
 601  HDIIKGNKEI TVGTFGVTAV SGHI
```

Figure 10

+ strand (sense)		sequence (5'-->3')
1. pch8-sp6-1f	369	GCT AAG CCA GAG CTA CAG G
2. pch8-sp6-2f	677	tCT GAT CTT CTG CTG ATT C (CTC)
3. pch8-1fa	1238	TCT GAA CTG CCT GAG AGA C
4. pch8-2f	1462	CCA AAT GGG AGC ATT ACA AG
5. pch8-3f	1745	TCA TCA AAT GAT CAG AAC C
6. pch8-4f	1995	ATT CTG GAG AGT TGG TAT CC
7. pch8-5f	2277	GGA ATA AGG AAA GAG CTT G
8. pch8-6f	2559	TCC ACT CAT ATT CCA ATA CC
9. pch8-5rb	2849	CCT GAG AGA CAG AAC TGT TC
10.pch8-4rb	3090	GGA CCC TTC ACT TCC TTA C
11.pch8-3rb	3370	GGC CAC CAC TTG TCC TGG G
12.pch8-2rb	3517	CAG AAC AGT GCT CTA ACT G
13.pch8-1rb	3970	GTA CTG CCT CTC TTA AAT G
- strand (antisense)		sequence (5'-->3')
14.pch8-2r	3617	CAG TTA CAG CAC TGT TCT G
15.pch8-3r	3360	CCC AGG ACA AGT GGT GGC C
16.pch8-4r	3140	GTA AGG AAG TGA AGG GTC C
17.pch8-5r	3849	GAA CAG TTC TGT CTC TCA GG
18.pch8-6r	3563	CTT GGG TAT TGG AAT ATG AG
19.pch8-5fb	2277	CAA GCT CTT TCC TTA TTC C
20.pch8-4fb	1999	ATA GGA TAC CAA CTC TCC AG
21.pch8-3fb	1746	TGG TTC TGA TCA TTT GAT G
22.pch8-2fb	1462	CTT GTA ATG CTC CCA TTT GG
23.pch8-1fb	1238	GTC TCT CAG GCA GTT CAG A
24.pch8-fb-1f	941	GTA GAG AAT CAC GTA CAG C
25.pch8-fb-2f	612	CAA TGA CCA GTA GCA TAA C
26.CH8-3670	3891	CAG CAT TTA AGA GAG GCA G
27.CH8a	387	CCT GTA GCT CTG GCT TAG CAT CC
28.CH8b	510	CCC CTT CAT TGA GAT CAT CTA G

Figure 11(A)

```
   1  GTGCGCCGTG GCGCGGCCCG GCTGACAGGT TCTTTAATGG AGGAGCCAAT
  51  CTCTCTGCAC ACCTGGTTTC ATCTAATAAT ATACAGACAC CAGCTCTGAG
 101  GCCAGTTAAT CATCCCCAGT GTCCAGGCAC AGAGTAGTCG GTCCGCCTCA
 151  CAATGTTGGA CTTTCTAGCC GAGAACAACC TCTGTGGCCA AGCAATCCTA
 201  AGGATTGTTT CCTGTGGTAA TGCCATCATT GCTGAACTTT TGAGACTCTC
 251  TGAGTTTATT CCTGCTGTGT TCAGGTTAAA AGACAGAGCT GATCAACAGA
 301  AATATGGAGA TATCATATTT GATTTCAGCT ATTTTAAGGG TCCAGAATTA
 351  TGGGAAAGCA AACTGGATGC TAAGCCAGAG CTACAGGATT TAGATGAAGA
 401  ATTTCGTGAA AACAACATAG AAATTGTGAC CAGATTTTAT TTAGCATTTC
 451  AAAGTGTACA TAAATATATT GTAGACTTAA ACAGATATCT AGATGATCTC
 501  AATGAAGGGG TTTATATTCA GCAAACCTTA GAAACTGTGC TTCTCAATGA
 551  AGATGGAAAA CAACTTCTAT GTGAAGCACT GTACTTATAT GGAGTTATGC
 601  TACTGGTCAT TGACCAAAAG ATTGAAGGAG AAGTCAGAGA GAGGATGCTG
 651  GTTTCTTACT ACCGATACAG TGCTGCTCGA TCTTCTGCTG ATTCAAATAT
 701  GGACGATATT TGTAAGCTGC TTCGAAGTAC AGGTTATTCT AGCCAACCAG
 751  GTGCCAAAAG ACCATCCAAC TATCCCGAGA GCTATTTCCA GAGAGTGCCT
 801  ATCAACGAAT CCTTCATCAG TATGGTCATT GGTCGACTGA GATCTGATGA
 851  TATTTACAAC CAGGTCTCAG CGTATCCTTT GCCGGAGCAT CGCAGCACAG
 901  CCCTGGCAAA CCAAGCTGCC ATGCTGTACG TGATTCTCTA CTTTGAGCCT
 951  TCCATCCTTC ACACCCATCA AGCAAAAATG AGAGAGATAG TGGATAAATA
1001  CTTTCCAGAT AATTGGGTAA TTAGTATTTA CATGGGGATC ACAGTTAATC
1051  TAGTAGATGC TTGGGAACCT TACAAAGCTG CAAAAACTGC TTTAAATAAT
1101  ACCCTGGACC TTTCAAATGT CAGAGAACAG GCAAGCAGAT ATGCTACTGT
1151  CAGTGAAAGA GTGCATGCTC AAGTGCAGCA ATTTCTAAAA GAAGGTTATT
1201  TAAGGGAGGA GATGGTTCTG GACAATATCC CAAAGCTTCT GAACTGCCTG
1251  AGAGACTGCA ATGTTGCCAT CCGATGGCTG ATGCTTCATA CAGCAGACTC
1301  AGCCTGTGAC CCAAACAACA AACGCCTTCG TCAAATCAAG GACCAGATTC
1351  TAACAGACTC TCGGTACAAT CCCAGGATCC TCTTCCAGCT GCTGTTAGAT
1401  ACTGCACAAT TTGAGTTTAT ACTCAAAGAG ATGTTCAAGC AAATGCTTTC
1451  AGAAAAGCAA ACCAAATGGG AGCATTACAA GAAAGAGGGT TCGGAGCGGA
1501  TGACTGAGCT TGCTGATGTC TTTTCAGGAG TGAAACCCCT AACCAGAGTG
1551  GAGAAAAATG AAAACCTTCA AGCTTGGTTC AGAGAGATCT CAAAACAAAT
1601  ATTGTCTTTA AATTATGATG ATTCTACTGC TGCGGGCAGA AAAACTGTAC
1651  AACTGATACA AGCTTTGGAA GAGGTTCAAG AATTCCACCA GTTGGAATCC
1701  AATCTGCAAG TATGTCAGTT TCTTGCCGAT ACTCGAAAGT TTCTTCATCA
1751  AATGATCAGA ACCATTAACA TTAAAGAGGA GGTTCTGATC ACAATGCAGA
1801  TCGTTGGGGA CCTTTCTTTC GCTTGGCAGT TGATTGACAG TTTCACATCC
1851  ATCATGCAAG AAAGCATAAG GGTAAATCCA TCCATGGTTA CTAAACTCAG
1901  AGCTACCTTC CTAAAGCTTG CCTCTGCCCT CGATCTGCCC CTTCTTCGTA
1951  TTAATCAGGC AAATCGCCCC GACCTGCTCA GCGTGTCACA GTACTATTCT
2001  GGAGAGTTGG TATCCTATGT GAGAAAAGTT TTGCAGATCA TCCCAGAAAG
2051  CATGTTTACA TCTCTTCTAA AGATCATAAA GCTTCAGACC CACGACATTA
2101  TTGAAGTGCC TACCCGCCTG GACAAAGACA AGCTGAGGGA CTATGCTCAG
2151  CTAGGCCCAC GATACGAGGT TGCCAAGCTT ACTCATGCTA TTTCCATTTT
2201  TACTGAAGGC ATCTTAATGA TGAAAACGAC TTTGGTTGGC ATCATCAAGG
2251  TGGATCCAAA GCAGTTGCTG GAAGATGGAA TAAGGAAAGA GCTTGTGAAG
2301  CGCGTTGCCT TTGCCCTGCA TAGGGGACTG ATATTCAACC CTCGAGCCAA
2351  GCCAAGTGAA TTGATGCCCA AGCTGAAAGA GTTGGGAGCG ACCATGGATG
2401  GATTCCATCG TTCTTTTGAA TACATACAGG ACTATGTCAA CATTTATGGT
2451  CTGAAGATTT GGCAGGAAGA AGTATCTCGT ATCATAAATT ACAACGTGGA
2501  GCAAGAGTGT AATAACTTTC TAAGAACGAA GATTCAAGAT TGGCAAAGCA
2551  TGTACCAGTC CACTCATATT CCAATACCCA AGTTTACCCC TGTGGATGAG
```

Figure 11(B)

```
2601    TCTGTAACGT TTATTGGTCG ACTCTGCAGA GAAATCCTGC GGATCACAGA
2651    CCCAAAAATG ACATGTCACA TAGACCAGCT GAACACTTGG TATGATATGA
2701    AAACTCATCA GGAAGTGACC AGCAGCCGCC TCTTCTCAGA AATCCAGACC
2751    ACCTTGGGAA CCTTTGGTCT AAATGGCTTA GACAGGCTTC TGTGCTTTAT
2801    GATTGTAAAA GAGTTACAGA ATTTCCTCAG TATGTTTCAG AAAATTATCC
2851    TGAGAGACAG AACTGTTCAG GACACTTTAA AAACCCTCAT GAATGCTGTC
2901    AGTCCCCTAA AAAGTATTGT CGCAAATTCA AATAAAATTT ATTTTTCCGC
2951    CATTGCCAAA ACACAGAAGA TTTGGACTGC GTATCTCGAG GCTATAATGA
3001    AGGTTGGGCA GATGCAGATT CTGAGGCAAC AGATTGCCAA TGAATTAAAT
3051    TATTCTTGTC GGTTTGATTC TAAACATCTG GCAGCTGCTC TGGAGAATCT
3101    CAATAAGGCT CTCCTAGCAG ACATTGAAGC CCACTATCAG GACCCTTCAC
3151    TTCCTTACCC CAAAGAAGAT AACACACTTT TATATGAAAT CACAGCCTAT
3201    CTGGAGGCAG CTGGCATTCA CAACCCACTG AATAAGATAT ACATAACAAC
3251    AAAGCGCTTA CCCTATTTTC CAATTGTAAA CTTTCTATTT TTGATCGCTC
3301    AGTTGCCAAA ACTTCAATAC AACAAAAATC TGGGAATGGT CTGCCGAAAA
3351    CCGACCGACC CGGTTGATTG GCCACCACTT GTCCTGGGAC TGCTCACTCT
3401    GCTGAAGCAG TTCCATTCCC GGTACACCGA GCAGCTCCTG GCGCTGATTG
3451    GCCAGTTTAT CTGCTCCACG GTGGAGCAGT GTACAAGCCA GAAGATACCT
3501    GAAATTCCTG CAGATGTTGT GGGTGCCCTT CTGTTCCTGG AGGATTATGT
3551    TCGGTACACA AAGCTACCCA GGAGGGTTGC TGAAGCACAT GTGCCTAATT
3601    TCATTTTTGA TGAGTTCAGA ACAGTGCTGT AACTGTTTTT CCTACTTCTT
3651    CAATGGAAGG ATTGTCCTTA GATCTTCCCA CCATCACAAA TGAATTTGAA
3701    GATGAAAAGA AACTCAGTTG CTCATACAAC TGCATTTTTT CTGTCTATTA
3751    TGGGAAACAT CAGACGTTAT GAGTAAGATA TATCTCATGG CATTAGTTAA
3801    TATAACTGAT ATTGTTTAAA TCATGGTATT ACATGCAATT TATATCAGAT
3851    AAAAGCAGAA CACATTTTTG TACTGCCTCT CTTAAATGCT GAATGTAACT
3901    GTTATGTATA AATCCATTTA GTTTTATGTT CTAAAGAACT ATTTGTGCAA
3951    CTCCAGATTT TCAGTAAAAT AGTATTACTA GT
```

Figure 12(A)

APWRGPADRF FNGGANLSAH LVSSNNIQTP ALRPVNHPQC PGTE•SVRLT
MLDFLAENNL CGQAILRIVS CGNAIIAELL RLSEFIPAVF RLKDRADQQK
YGDIIFDFSY FKGPELWESK LDAKPELQDL DEEFRENNIE IVTRFYLAFQ
SVHKYIVDLN RYLDDLNEGV YIQQTLETVL LNEDGKQLLC EALYLYGVML
LVIDQKIEGE VRERMLVSYY RYSAARSSAD SNMDDICKLL RSTGYSSQPG
AKRPSNYPES YFQRVPINES FISMVIGRLR SDDIYNQVSA YPLPEHRSTA
LANQAAMLYV ILYFEPSILH THQAKMREIV DKYFPDNWVI SIYMGITVNL
VDAWEPYKAA KTALNNTLDL SNVREQASRY ATVSERVHAQ VQQFLKEGYL
REEMVLDNIP KLLNCLRDCN VAIRWLMLHT ADSACDPNNK RLRQIKDQIL
TDSRYNPRIL FQLLLDTAQF EFILKEMFKQ MLSEKQTKWE HYKKEGSERM
TELADVFSGV KPLTRVEKNE NLQAWFREIS KQILSLNYDD STAAGRKTVQ
LIQALEEVQE FHQLESNLQV CQFLADTRKF LHQMIRTINI KEEVLITMQI
VGDLSFAWQL IDSFTSIMQE SIRVNPSMVT KLRATFLKLA SALDLPLLRI
NQANRPDLLS VSQYYSGELV SYVRKVLQII PESMFTSLLK IIKLQTHDII
EVPTRLDKDK LRDYAQLGPR YEVAKLTHAI SIFTEGILMM KTTLVGIIKV
DPKQLLEDGI RKELVKRVAF ALHRGLIFNP RAKPSELMPK LKELGATMDG
FHRSFEYIQD YVNIYGLKIW QEEVSRIINY NVEQECNNFL RTKIQDWQSM
YQSTHIPIPK FTPVDESVTF IGRLCREILR ITDPKMTCHI DQLNTWYDMK
THQEVTSSRL FSEIQTTLGT FGLNGLDRLL CFMIVKELQN FLSMFQKIIL
RDRTVQDTLK TLMNAVSPLK SIVANSNKIY FSAIAKTQKI WTAYLEAIMK
VGQMQILRQQ IANELNYSCR FDSKHLAAAL ENLNKALLAD IEAHYQDPSL
PYPKEDNTLL YEITAYLEAA GIHNPLNKIY ITTKRLPYFP IVNFLFLIAQ
LPKLQYNKNL GMVCRKPTDP VDWPPLVLGL LTLLKQFHSR YTEQLLALIG
QFICSTVEQC TSQKIPEIPA DVVGALLFLE DYVRYTKLPR RVAEAHVPNF
IFDEFRTVL• LFFLLLQWKD CP•IFPPSQM NLKMKRNSVA HTTAFFLSIM
GNIRRYE•DI SHGIS•YN•Y CLNHGITCNL YQIKAEHIFV LPLLNAECNC
YV•IHLVLCS KELFVQLQIF SKIVLL

Figure 12(B)

MLDFLAENNL CGQAILRIVS CGNAIIAELL RLSEFIPAVF RLKDRADQQK
YGDIIFDFSY FKGPELWESK LDAKPELQDL DEEFRENNIE IVTRFYLAFQ
SVHKYIVDLN RYLDDLNEGV YIQQTLETVL LNEDGKQLLC EALYLYGVML
LVIDQKIEGE VRERMLVSYY RYSAARSSAD SNMDDICKLL RSTGYSSQPG
AKRPSNYPES YFQRVPINES FISMVIGRLR SDDIYNQVSA YPLPEHRSTA
LANQAAMLYV ILYFEPSILH THQAKMREIV DKYFPDNWVI SIYMGITVNL
VDAWEPYKAA KTALNNTLDL SNVREQASRY ATVSERVHAQ VQQFLKEGYL
REEMVLDNIP KLLNCLRDCN VAIRWLMLHT ADSACDPNNK RLRQIKDQIL
TDSRYNPRIL FQLLLDTAQF EFILKEMFKQ MLSEKQTKWE HYKKEGSERM
TELADVFSGV KPLTRVEKNE NLQAWFREIS KQILSLNYDD STAAGRKTVQ
LIQALEEVQE FHQLESNLQV CQFLADTRKF LHQMIRTINI KEEVLITMQI
VGDLSFAWQL IDSFTSIMQE SIRVNPSMVT KLRATFLKLA SALDLPLLRI
NQANRPDLLS VSQYYSGELV SYVRKVLQII PESMFTSLLK IIKLQTHDII
EVPTRLDKDK LRDYAQLGPR YEVAKLTHAI SIFTEGILMM KTTLVGIIKV
DPKQLLEDGI RKELVKRVAF ALHRGLIFNP RAKPSELMPK LKELGATMDG
FHRSFEYIQD YVNIYGLKIW QEEVSRIINY NVEQECNNFL RTKIQDWQSM
YQSTHIPIPK FTPVDESVTF IGRLCREILR ITDPKMTCHI DQLNIWYDMK
THQEVTSSRL FSEIQTTLGT FGLNGLDRLL CFMIVKELQN FLSMFQKIIL
RDRTVQDTLK TLMNAVSPLK SIVANSNKIY FSAIAKTQKI WTAYLEAIMK
VGQMQILRQQ IANELNYSCR FDSKHLAAAL ENLNKALLAD IEAHYQDPSL
PYPKEDNTLL YEITAYLEAA GIHNPLNKIY ITTKRLPYFP IVNFLFLIAQ
LPKLQYNKNL GMVCRKPTDP VDWPPLVLGL LTLLKQFHSR YTEQLLALIG
QFICSTVEQC TSQKIPEIPA DVVGALLFLE DYVRYTKLPR RVAEAHVPNF
IFDEFRTVL

Figure 13(A)

AGG GGC GGA AGT CGG GGT CTG ACC CGC TCC AGG TCC GGG ACT GCG GAT
AGA AGA GGA CCG CCG CCT TGA GGG AGG GGT GGA AAC TGG GTG CCG GCT
CCG CGC GCG ACC TCC GGC CCT GCG CGT GCG CCG TGG CGC GGC CCG GCT
GAC AGG TTC TTT AAT GGA GGA GCC AAT CTC TCT GCA CAC CTG GTT TCA
TCT AAT AAT ATA CAG ACA CCA GCT CTG AGG CCA GTT AAT CAT CCC CAG
TGT CCA GGC ACA GAG TAG TCG GTC CGC CTC ACA ATG TTG GAC TTT CTA
GCC GAG AAC AAC CTC TGT GGC CAA GCA ATC CTA AGG ATT GTT TCC TGT
GGT AAT GCC ATC ATT GCT GAA CTT TTG AGA CTC TCT GAG TTT ATT CCT
GCT GTG TTC AGG TTA AAA GAC AGA GCT GAT CAA CAG AAA TAT GGA GAT
ATC ATA TTT GAT TTC AGC TAT TTT AAG GGT CCA GAA TTA TGG GAA AGC
AAA CTG GAT GCT AAG CCA GAG CTA CAG GAT TTA GAT GAA GAA TTT CGT
GAA AAC AAC ATA GAA ATT GTG ACC AGA TTT TAT TTA GCA TTT CAA AGT
GTA CAT AAA TAT ATT GTA GAC TTA AAC AGA TAT CTA GAT GAT CTC AAT
GAA GGG GTT TAT ATT CAG CAA ACC TTA GAA ACT GTG CTT CTC AAT GAA
GAT GGA AAA CAA CTT CTA TGT GAA GCA CTG TAC TTA TAT GGA GTT ATG
CTA CTG GTC ATT GAC CAA AAG ATT GAA GGA GAA GTC AGA GAG AGG ATG
CTG GTT TCT TAC TAC CGA TAC AGT GCT GCT CGA TCT TCT GCT GAT TCA
AAT ATG GAC GAT ATT TGT AAG CTG CTT CGA AGT ACA GGT TAT TCT AGC
CAA CCA GGT GCC AAA AGA CCA TCC AAC TAT CCC GAG AGC TAT TTC CAG
AGA GTG CCT ATC AAC GAA TCC TTC ATC AGT ATG GTC ATT GGT CGA CTG
AGA TCT GAT GAT ATT TAC AAC CAG GTC TCA GCG TAT CCT TTG CCG GAG
CAT CGC AGC ACA GCC CTG GCA AAC CAA GCT GCC ATG CTG TAC GTG ATT
CTC TAC TTT GAG CCT TCC ATC CTT CAC ACC CAT CAA GCA AAA ATG AGA
GAG ATA GTG GAT AAA TAC TTT CCA GAT AAT TGG GTA ATT AGT ATT TAC
ATG GGG ATC ACA GTT AAT CTA GTA GAT GCT TGG GAA CCT TAC AAA GCT
GCA AAA ACT GCT TTA AAT AAT ACC CTG GAC CTT TCA AAT GTC AGA GAA
CAG GCA AGC AGA TAT GCT ACT GTC AGT GAA AGA GTG CAT GCT CAA GTG
CAG CAA TTT CTA AAA GAA GGT TAT TTA AGG GAG GAG ATG GTT CTG GAC
AAT ATC CCA AAG CTT CTG AAC TGC CTG AGA GAC TGC AAT GTT GCC ATC
CGA TGG CTG ATG CTT CAT ACA GCA GAC TCA GCC TGT GAC CCA AAC AAC
AAA CGC CTT CGT CAA ATC AAG GAC CAG ATT CTA ACA GAC TCT CGG TAC
AAT CCC AGG ATC CTC TTC CAG CTG CTG TTA GAT ACT GCA CAA TTT GAG
TTT ATA CTC AAA GAG ATG TTC AAG CAA ATG CTT TCA GAA AAG CAA ACC
AAA TGG GAG CAT TAC AAG AAA GAG GGT TCG GAG CGG ATG ACT GAG CTT
GCT GAT GTC TTT TCA GGA GTG AAA CCC CTA ACC AGA GTG GAG AAA AAT
GAA AAC CTT CAA GCT TGG TTC AGA GAG ATC TCA AAA CAA ATA TTG TCT
TTA AAT TAT GAT GAT TCT ACT GCT GCG GGC AGA AAA ACT GTA CAA CTG
ATA CAA GCT TTG GAA GAG GTT CAA GAA TTC CAC CAG TTG GAA TCC AAT
CTG CAA GTA TGT CAG TTT CTT GCC GAT ACT CGA AAG TTT CTT CAT CAA
ATG ATC AGA ACC ATT AAC ATT AAA GAG GAG GTT CTG ATC ACA ATG CAG
ATC GTT GGG GAC CTT TCT TTC GCT TGG CAG TTG ATT GAC AGT TTC ACA
TCC ATC ATG CAA GAA AGC ATA AGG GTA AAT CCA TCC ATG GTT ACT AAA
CTC AGA GCT ACC TTC CTA AAG CTT GCC TCT GCC CTC GAT CTG CCC CTT
CTT CGT ATT AAT CAG GCA AAT CGC CCC GAC CTG CTC AGC GTG TCA CAG
TAC TAT TCT GGA GAG TTG GTA TCC TAT GTG AGA AAA GTT TTG CAG ATC
ATC CCA GAA AGC ATG TTT ACA TCT CTT CTA AAG ATC ATA AAG CTT CAG
ACC CAC GAC ATT ATT GAA GTG CCT ACC CGC CTG GAC AAA GAC AAG CTG
AGG GAC TAT GCT CAG CTA GGC CCA CGA TAC GAG GTT GCC AAG CTT ACT
CAT GCT ATT TCC ATT TTT ACT GAA GGC ATC TTA ATG ATG AAA ACG ACT

Figure 13(B)

TTG GTT GGC ATC ATC AAG GTG GAT CCA AAG CAG TTG CTG GAA GAT GGA
ATA AGG AAA GAG CTT GTG AAG CGC GTT GCC TTT GCC CTG CAT AGG GGA
CTG ATA TTC AAC CCT CGA GCC AAG CCA AGT GAA TTG ATG CCC AAG CTG
AAA GAG TTG GGA GCG ACC ATG GAT GGA TTC CAT CGT TCT TTT GAA TAC
ATA CAG GAC TAT GTC AAC ATT TAT GGT CTG AAG ATT TGG CAG GAA GAA
GTA TCT CGT ATC ATA AAT TAC AAC GTG GAG CAA GAG TGT AAT AAC TTT
CTA AGA ACG AAG ATT CAA GAT TGG CAA AGC ATG TAC CAG TCC ACT CAT
ATT CCA ATA CCC AAG TTT ACC CCT GTG GAT GAG TCT GTA ACG TTT ATT
GGT CGA CTC TGC AGA GAA ATC CTG CGG ATC ACA GAC CCA AAA ATG ACA
TGT CAC ATA GAC CAG CTG AAC ACT TGG TAT GAT ATG AAA ACT CAT CAG
GAA GTG ACC AGC AGC CGC CTC TTC TCA GAA ATC CAG ACC ACC TTG GGA
ACC TTT GGT CTA AAT GGC TTA GAC AGG CTT CTG TGC TTT ATG ATT GTA
AAA GAG TTA CAG AAT TTC CTC AGT ATG TTT CAG AAA ATT ATC CTG AGA
GAC AGA ACT GTT CAG GAC ACT TTA AAA ACC CTC ATG AAT GCT GTC AGT
CCC CTA AAA AGT ATT GTC GCA AAT TCA AAT AAA ATT TAT TTT TCC GCC
ATT GCC AAA ACA CAG AAG ATT TGG ACT GCG TAT CTC GAG GCT ATA ATG
AAG GTT GGG CAG ATG CAG ATT CTG AGG CAA CAG ATT GCC AAT GAA TTA
AAT TAT TCT TGT CGG TTT GAT TCT AAA CAT CTG GCA GCT GCT CTG GAG
AAT CTC AAT AAG GCT CTC CTA GCA GAC ATT GAA GCC CAC TAT CAG GAC
CCT TCA CTT CCT TAC CCC AAA GAA GAT AAC ACA CTT TTA TAT GAA ATC
ACA GCC TAT CTG GAG GCA GCT GGC ATT CAC AAC CCA CTG AAT AAG ATA
TAC ATA ACA ACA AAG CGC TTA CCC TAT TTT CCA ATT GTA AAC TTT CTA
TTT TTG ATC GCT CAG TTG CCA AAA CTT CAA TAC AAC AAA AAT CTG GGA
ATG GTC TGC CGA AAA CCG ACC GAC CCG GTT GAT TGG CCA CCA CTT GTC
CTG GGA CTG CTC ACT CTG CTG AAG CAG TTC CAT TCC CGG TAC ACC GAG
CAG CTC CTG GCG CTG ATT GGC CAG TTT ATC TGC TCC ACG GTG GAG CAG
TGT ACA AGC CAG AAG ATA CCT GAA ATT CCT GCA GAT GTT GTG GGT GCC
CTT CTG TTC CTG GAG GAT TAT GTT CGG TAC ACA AAG CTA CCC AGG AGG
GTT GCT GAA GCA CAT GTG CCT AAT TTC ATT TTT GAT GAG TTC AGA ACA
GTG CTG TAA CTG TTT TTC CTA CTT CTT CAA TGG AAG GAT TGT CCT TAG
ATC TTC CCA CCA TCA CAA ATG AAT TTG AAG ATG AAA AGA AAC TCA GTT
GCT CAT ACA ACT GCA TTT TTT CTG TCT ATT ATG GGA AAC ATC AGA CGT
TAT GAG TAA GAT ATA TCT CAT GGC ATT AGT TAA TAT AAC TGA TAT TGT
TTA AAT CAT GGT ATT ACA TGC AAT TTA TAT CAG ATA AAA GCA GAA CAC
ATT TTT GTA CTG CCT CTC TTA AAT GCT GAA TGT AAC TGT TAT GTA TAA
ATC CAT TTA GTT TTA TGT TCT AAA GAA CTA TTT GTG CAA CTC CAG ATT
TTC AGT AAA ATA GTA TTA CTA GT

Figure 14(A)

Arg Gly Gly Ser Arg Gly Leu Thr Arg Ser Arg Ser Gly Thr Ala Asp
Arg Arg Gly Pro Pro Pro * Gly Arg Gly Gly Asn Trp Val Pro Ala
Pro Arg Ala Thr Ser Gly Pro Ala Arg Ala Pro Trp Arg Gly Pro Ala
Asp Arg Phe Phe Asn Gly Gly Ala Asn Leu Ser Ala His Leu Val Ser
Ser Asn Asn Ile Gln Thr Pro Ala Leu Arg Pro Val Asn His Pro Gln
Cys Pro Gly Thr Glu * Ser Val Arg Leu Thr Met Leu Asp Phe Leu
Ala Glu Asn Asn Leu Cys Gly Gln Ala Ile Leu Arg Ile Val Ser Cys
Gly Asn Ala Ile Ile Ala Glu Leu Leu Arg Leu Ser Glu Phe Ile Pro
Ala Val Phe Arg Leu Lys Asp Arg Ala Asp Gln Gln Lys Tyr Gly Asp
Ile Ile Phe Asp Phe Ser Tyr Phe Lys Gly Pro Glu Leu Trp Glu Ser
Lys Leu Asp Ala Lys Pro Glu Leu Gln Asp Leu Asp Glu Glu Phe Arg
Glu Asn Asn Ile Glu Ile Val Thr Arg Phe Tyr Leu Ala Phe Gln Ser
Val His Lys Tyr Ile Val Asp Leu Asn Arg Tyr Leu Asp Asp Leu Asn
Glu Gly Val Tyr Ile Gln Gln Thr Leu Glu Thr Val Leu Leu Asn Glu
Asp Gly Lys Gln Leu Leu Cys Glu Ala Leu Tyr Leu Tyr Gly Val Met
Leu Leu Val Ile Asp Gln Lys Ile Glu Gly Glu Val Arg Glu Arg Met
Leu Val Ser Tyr Tyr Arg Tyr Ser Ala Ala Arg Ser Ser Ala Asp Ser
Asn Met Asp Asp Ile Cys Lys Leu Leu Arg Ser Thr Gly Tyr Ser Ser
Gln Pro Gly Ala Lys Arg Pro Ser Asn Tyr Pro Glu Ser Tyr Phe Gln
Arg Val Pro Ile Asn Glu Ser Phe Ile Ser Met Val Ile Gly Arg Leu
Arg Ser Asp Asp Ile Tyr Asn Gln Val Ser Ala Tyr Pro Leu Pro Glu
His Arg Ser Thr Ala Leu Ala Asn Gln Ala Ala Met Leu Tyr Val Ile
Leu Tyr Phe Glu Pro Ser Ile Leu His Thr His Gln Ala Lys Met Arg
Glu Ile Val Asp Lys Tyr Phe Pro Asp Asn Trp Val Ile Ser Ile Tyr
Met Gly Ile Thr Val Asn Leu Val Asp Ala Trp Glu Pro Tyr Lys Ala
Ala Lys Thr Ala Leu Asn Asn Thr Leu Asp Leu Ser Asn Val Arg Glu
Gln Ala Ser Arg Tyr Ala Thr Val Ser Glu Arg Val His Ala Gln Val
Gln Gln Phe Leu Lys Glu Gly Tyr Leu Arg Glu Glu Met Val Leu Asp
Asn Ile Pro Lys Leu Leu Asn Cys Leu Arg Asp Cys Asn Val Ala Ile
Arg Trp Leu Met Leu His Thr Ala Asp Ser Ala Cys Asp Pro Asn Asn
Lys Arg Leu Arg Gln Ile Lys Asp Gln Ile Leu Thr Asp Ser Arg Tyr
Asn Pro Arg Ile Leu Phe Gln Leu Leu Leu Asp Thr Ala Gln Phe Glu
Phe Ile Leu Lys Glu Met Phe Lys Gln Met Leu Ser Glu Lys Gln Thr
Lys Trp Glu His Tyr Lys Lys Glu Gly Ser Glu Arg Met Thr Glu Leu
Ala Asp Val Phe Ser Gly Val Lys Pro Leu Thr Arg Val Glu Lys Asn
Glu Asn Leu Gln Ala Trp Phe Arg Glu Ile Ser Lys Gln Ile Leu Ser
Leu Asn Tyr Asp Asp Ser Thr Ala Ala Gly Arg Lys Thr Val Gln Leu
Ile Gln Ala Leu Glu Glu Val Gln Glu Phe His Gln Leu Glu Ser Asn
Leu Gln Val Cys Gln Phe Leu Ala Asp Thr Arg Lys Phe Leu His Gln
Met Ile Arg Thr Ile Asn Ile Lys Glu Glu Val Leu Ile Thr Met Gln
Ile Val Gly Asp Leu Ser Phe Ala Trp Gln Leu Ile Asp Ser Phe Thr
Ser Ile Met Gln Glu Ser Ile Arg Val Asn Pro Ser Met Val Thr Lys
Leu Arg Ala Thr Phe Leu Lys Leu Ala Ser Ala Leu Asp Leu Pro Leu
Leu Arg Ile Asn Gln Ala Asn Arg Pro Asp Leu Leu Ser Val Ser Gln
Tyr Tyr Ser Gly Glu Leu Val Ser Tyr Val Arg Lys Val Leu Gln Ile
Ile Pro Glu Ser Met Phe Thr Ser Leu Leu Lys Ile Ile Lys Leu Gln
Thr His Asp Ile Ile Glu Val Pro Thr Arg Leu Asp Lys Asp Lys Leu
Arg Asp Tyr Ala Gln Leu Gly Pro Arg Tyr Glu Val Ala Lys Leu Thr
His Ala Ile Ser Ile Phe Thr Glu Gly Ile Leu Met Met Lys Thr Thr
Leu Val Gly Ile Ile Lys Val Asp Pro Lys Gln Leu Leu Glu Asp Gly
Ile Arg Lys Glu Leu Val Lys Arg Val Ala Phe Ala Leu His Arg Gly

Figure 14(B)

Leu Ile Phe Asn Pro Arg Ala Lys Pro Ser Glu Leu Met Pro Lys Leu
Lys Glu Leu Gly Ala Thr Met Asp Gly Phe His Arg Ser Phe Glu Tyr
Ile Gln Asp Tyr Val Asn Ile Tyr Gly Leu Lys Ile Trp Gln Glu Glu
Val Ser Arg Ile Ile Asn Tyr Asn Val Glu Gln Glu Cys Asn Asn Phe
Leu Arg Thr Lys Ile Gln Asp Trp Gln Ser Met Tyr Gln Ser Thr His
Ile Pro Ile Pro Lys Phe Thr Pro Val Asp Glu Ser Val Thr Phe Ile
Gly Arg Leu Cys Arg Glu Ile Leu Arg Ile Thr Asp Pro Lys Met Thr
Cys His Ile Asp Gln Leu Asn Thr Trp Tyr Asp Met Lys Thr His Gln
Glu Val Thr Ser Ser Arg Leu Phe Ser Glu Ile Gln Thr Thr Leu Gly
Thr Phe Gly Leu Asn Gly Leu Asp Arg Leu Leu Cys Phe Met Ile Val
Lys Glu Leu Gln Asn Phe Leu Ser Met Phe Gln Lys Ile Ile Leu Arg
Asp Arg Thr Val Gln Asp Thr Leu Lys Thr Leu Met Asn Ala Val Ser
Pro Leu Lys Ser Ile Val Ala Asn Ser Asn Lys Ile Tyr Phe Ser Ala
Ile Ala Lys Thr Gln Lys Ile Trp Thr Ala Tyr Leu Glu Ala Ile Met
Lys Val Gly Gln Met Gln Ile Leu Arg Gln Gln Ile Ala Asn Glu Leu
Asn Tyr Ser Cys Arg Phe Asp Ser Lys His Leu Ala Ala Ala Leu Glu
Asn Leu Asn Lys Ala Leu Leu Ala Asp Ile Glu Ala His Tyr Gln Asp
Pro Ser Leu Pro Tyr Pro Lys Glu Asp Asn Thr Leu Leu Tyr Glu Ile
Thr Ala Tyr Leu Glu Ala Ala Gly Ile His Asn Pro Leu Asn Lys Ile
Tyr Ile Thr Thr Lys Arg Leu Pro Tyr Phe Pro Ile Val Asn Phe Leu
Phe Leu Ile Ala Gln Leu Pro Lys Leu Gln Tyr Asn Lys Asn Leu Gly
Met Val Cys Arg Lys Pro Thr Asp Pro Val Asp Trp Pro Pro Leu Val
Leu Gly Leu Leu Thr Leu Leu Lys Gln Phe His Ser Arg Tyr Thr Glu
Gln Leu Leu Ala Leu Ile Gly Gln Phe Ile Cys Ser Thr Val Glu Gln
Cys Thr Ser Gln Lys Ile Pro Glu Ile Pro Ala Asp Val Val Gly Ala
Leu Leu Phe Leu Glu Asp Tyr Val Arg Tyr Thr Lys Leu Pro Arg Arg
Val Ala Glu Ala His Val Pro Asn Phe Ile Phe Asp Glu Phe Arg Thr
Val Leu * Leu Phe Phe Leu Leu Leu Gln Trp Lys Asp Cys Pro *
Ile Phe Pro Pro Ser Gln Met Asn Leu Lys Met Lys Arg Asn Ser Val
Ala His Thr Thr Ala Phe Phe Leu Ser Ile Met Gly Asn Ile Arg Arg
Tyr Glu * Asp Ile Ser His Gly Ile Ser * Tyr Asn * Tyr Cys
Leu Asn His Gly Ile Thr Cys Asn Leu Tyr Gln Ile Lys Ala Glu His
Ile Phe Val Leu Pro Leu Leu Asn Ala Glu Cys Asn Cys Tyr Val *
Ile His Leu Val Leu Cys Ser Lys Glu Leu Phe Val Gln Leu Gln Ile
Phe Ser Lys Ile Val Leu Leu

Figure 15

+ strand (sense)	1st base	sequence (5'-->3')
1. pch13-sp6-1f	370	TTT ACT TCT AAC GCT TAT TC
2. pch13-sp6-2f	726	TGA AGG AGT CCT TTG AGA CG
3. T7.1	1140	TCA CAA TGG GCT ACT GG
4. T7.2	1361	TTC AAC GAG GGA GAT GG
5. T7.3	1602	TTA GCA CCA CTG AGA GA
6. T7.4	2041	GTT CTT TTA GGC ATT TA
7. ch13-2480	2486	GCT GCG TCT GTT CGT CAG C
- strand (antisense)		
8. SP6.1	2746	CCT CTG CTT CAC AAC AT
9. SP6.2	2490	GCA GtA GGG CGG ACA CC
10. SP6.3	2213	AGG GTC TTC TTC ATT GT (C)
11. SP6.4	1812	GGA TTG TCT TTG TCT CT
12.pch13-t7-1f	1165	AGT GCA CTT CCA TGG GCG TG
13.pch13-t7-1fa	712	CCT TCA TCA GGT TGA CGA AC
14.pch13-t7-2fa	286	GCG GCA ATC AGA AAC GGA AG
15.CH13-AS-1	536	TGA ACA CGT GGT ACA T

Figure 16(A)

```
   1  CTTCCCTGAG CCCTTTCTGC CTGTGTAGGA AGCAGAAGGC GGAATGTCGG
  51  CTCTGCCCTT CTCCGTAAGA TGGTGCATTA AAACGTTCCT TATAAACTGG
 101  AAATGAAGGC TTGGGAAGAT GGCTAAAATC AGCAATCCTT GGAATAACGC
 151  AGAAGCATCC CTGCTTCCCT GGGCCCGCCC GTGGGCCTGC TTGTGCTGTT
 201  CAGTAGGTGG TTTTTAGAAA GGGCTTCCTT CAGCGTCATT AGCAACAGGA
 251  GTCGTCGTCC GTTTGCATGA GGAAATGTTC TTAACCTTCC GTTTCTGATT
 301  GCCTCTAGAC TGCATCTGTC ATAGACAAAT GCCCCCATCT TTTACAGAGA
 351  ACCAGTCTCT TCTTTAAACT TTACTTCTAA CGCTTATTCT TTTTACCTTA
 401  TATAGGAAAC CACTGATTGC TTGTGTGGAG AAACAGCTAT TAGGAGAACA
 451  TTTAACAGCA ATTCTGCAGA AAGGGCTCGA CCACTTACTG GATGAGAACA
 501  GAGTGCCGGA CCTCGCACAG ATGTACCAGC TGTTCAGCCG GGTGAGGGGC
 551  GGGCAGCAGG CGCTGCTGCA GCACTGGAGC GAGTACATCA AGACTTTTGG
 601  AACAGCGATC GTAATCAATC CTGAGAAAGA CAAAGACATG GTCCAAGACC
 651  TGTTGGACTT CAAGGACAAG GTGGACCACG TGATCGAGGT CTGCTTCCAG
 701  AAGAATGAGC GGTTCGTCAA CCTGATGAAG GAGTCCTTTG AGACGTTCAT
 751  CAACAAGAGA CCCAACAAGC CTGCAGAACT GATCGCAAAG CATGTGGATT
 801  CAAAGTTAAG AGCAGGCAAC AAAGAAGCCA CAGACGAGGA GCTGGAGCGG
 851  ACGTTGGACA AGATCATGAT CCTGTTCAGG TTTATCCACG GTAAAGATGT
 901  CTTTGAAGCA TTTTATAAAA AAGATTTGGC AAAAAGACTC CTTGTTGGGA
 951  AAAGTGCCTC AGTCGATGCT GAAAAGTCTA TGTTGTCAAA GCTCAAGCAT
1001  GAGTGCGGTG CAGCCTTCAC CAGCAAGCTG GAAGGCATGT TCAAGGACAT
1051  GGAGCTTTCG AAGGACATCA TGGTTCATTT CAAGCAGCAT ATGCAGAATC
1101  AGAGTGACTC AGGCCCTATA GACCTCACAG TGAACATACT CACAATGGGC
1151  TACTGGCCAA CATACACGCC CATGGAAGTG CACTTAACCC CAGAAATGAT
1201  TAAACTTCAG GAAGTATTTA AGGCATTTTA TCTTGGAAAG CACAGTGGTC
1251  GAAAACTTCA GTGGCAAACT ACTTTGGGAC ATGCTGTTTT AAAAGCGGAG
1301  TTTAAAGAAG GGAAGAAGGA ATTCCAGGTG TCCCTCTTCC AGACACTGGT
1351  GCTCCTCATG TTCAACGAGG GAGATGGCTT CAGCTTTGAG GAGATAAAAA
1401  TGGCCACGGG GATAGAGGAT AGTGAATTGC GCAGAACGCT GCAGTCCCTG
1451  GCCTGTGGCA AAGCACGTGT GCTGATTAAA AGTCCCAAAG GAAAGGAAGT
1501  GGAAGATGGA GACAAGTTCA TTTTTAATGG AGAGTTCAAG CACAAGTTGT
1551  TTAGAATAAA GATCAATCAA ATTCAGATGA AGGAAACTGT TGAGGAACAG
1601  GTTAGCACCA CTGAGAGAGT GTTTCAGGAT AGACAATATC AGATTGATGC
1651  TGCTATCGTC AGAATAATGA AGATGAGAAA GACTCTTGGT CATAATCTTC
1701  TAGTTTCTGA ATTATATAAT CAGCTGAAAT TTCCAGTAAA GCCTGGAGAT
1751  TTGAAAAAGA GAATTGAATC TCTGATAGAC AGAGACTATA TGGAGAGAGA
1801  CAAAGACAAT CCGAATCAGT ACCACTACGT GGCCTGACGC ATCTGCAGAC
1851  GGTTCCCCTT CATGAAACAC TAGAATGTAC CCTCAGAGCA GGAAGCACAC
1901  CTGTGCCATT TCTGGGACTC TGATTGATCC AGCTGTGGAC ATTGGAAGGC
1951  GAAGGAAGGG AGGTGGCTCC TGGGTCATCT TTCACAAGGC TCAAGACTTC
2001  AACCTGCAGA TGTATCTTTT TCCCTCCAGT TTTTCCTCTA GTTCTTTTAG
2051  GCATTTAAAT TGTTTCTGTT ACTCTGTGCA AAATAACTTT GAGATTGGAC
2101  AAGAAGATGT TACTAAAGAG AAGTTCCTTT AAAAGGTCTT GTTCTTGTGT
2151  CAAAAAGCTG CAAGTTTGGT TTGTTCTCGT GTGTGATCAT GAGTGCACAA
2201  TGAAGAAGAC CCTAGATGCT GCATTTTTTA GCTCTGAAGA TTCCTTAGGT
2251  ATCCCTGAAG ACAGCTCGCT CAGATGATCA GCATTTAGAG TGAAAACAAG
2301  GGCCCTTCAT GGGTGAACAT TAGAAAGAGC CAGGGTTCAA AGCTGGCGAA
2351  TGGATGACGC ACCCTAGCCA CTGGCCCCTC CCTGTTTCAT GTATTTCCAA
2401  AAGTTGTAAA CTTTGGTGGC TGATTTTTCG TAAGTCAGGT TTCTAAGTGA
2451  GCTCCCTGAG GTGCCAAGGC CATGGTGTCC GCCCTGCTGC GTCTGTTCGT
2501  CAGCTGAGTT CCTTGTGAAT CTCTGTTTTA GGGGTTGGGG CTAGTGTGTT
```

Figure 16(B)

```
2551  TGTGTTTCCA TTCTAAGATT GAGTCTGGCA GTCCCTGTTT TTTTGCATTG
2601  GGGTAACTGC TCTTTGATTT TTTTTAATTG CAGTATTTGT GTGATTGCAA
2651  TAATAAAGTT TGGTTTGGTT TTTACAGTCA TGCGCAGGGA CGATCCTTGT
2701  TCTCTGCTGT AAACTGTAAA AAGTTTATGG AGACCTAAAG TCTTGATGTT
2751  GTGAAGCAGA GGTTATTTTG TGGAAAGATT AAAAGGATTT TGTTGGTACC
2801  TGGTTTTGTG TTGTGTATAT ATACATGAGG TTGAACAGTG AAAGGAAAGT
2851  TCAGTAGTGA TGTTAGAAGG GTAACTATGA CAAAGATACT TTTGAGATAA
2901  CATTTAAAAG TACTTTATAT TTTACATAAT AGCATGTTTC ATTTTGATTA
2951  AAAGCTACCA AAGGAATTTT GATCATGGCA TAAGTGTTTA AAGCAATATT
3001  TTCTGGAATA TACCAAGTTT ATATAATTTG ATTTTGTGCT AAATTATTAA
3051  GAGTCTCTTT TTGAAACATG CGGGTTTGAA ATATGACACC TTGTGGGTTT
3101  CCATATTAAA ATCCTCACTC TTTAATTGTC ATTTTTATCT TTGAAAATTT
3151  TCATTTATGA GTTCCATGAT ATGTGGTCTA AGAAAGACCA AACAGATTTC
3201  TATTTTTTTT TCTTATAAGT TCGTTGTGTC TAGAGATTGT TAATATTGTA
3251  ATTTAATGTA GACTTACTTT GAATAAAATT AGTTTAATTG GCCTTAAAAT
3301  TACATTAATA AAACTTTGTG ATATGCAAAT GACACATTC
```

Figure 17

```
   1    FPEPFLPV•E AEGGMSALPF SVRWCIKTFL INWK•RLGKM AKISNPWNNA
  51    EASLLPWARP WACLCCSVGG F•KGLPSASL ATGVVVRLHE EMFLTFRF•L
 101    PLDCICHRQM PPSFTENQSL L•TLLLTLIL FTLYRKPLIA CVEKQLLGEH
 151    LTAILQKGLD HLLDENRVPD LAQMYQLFSR VRGGQQALLQ HWSEYIKTFG
 201    TAIVINPEKD KDMVQDLLDF KDKVDHVIEV CFQKNERFVN LMKESFETFI
 251    NKRPNKPAEL IAKHVDSKLR AGNKEATDEE LERTLDKIMI LFRFIHGKDV
 301    FEAFYKKDLA KRLLVGKSAS VDAEKSMLSK LKHECGAAFT SKLEGMFKDM
 351    ELSKDIMVHF KQHMQNQSDS GPIDLTVNIL TMGYWPTYTP MEVHLTPEMI
 401    KLQEVFKAFY LGKHSGRKLQ WQTTLGHAVL KAEFKEGKKE FQVSLFQTLV
 451    LLMFNEGDGF SFEEIKMATG IEDSELRRTL QSLACGKARV LIKSPKGKEV
 501    EDGDKFIFNG EFKHKLFRIK INQIQMKETV EEQVSTTERV FQDRQYQIDA
 551    AIVRIMKMRK TLGHNLLVSE LYNQLKFPVK PGDLKKRIES LIDRDYMERD
 601    KDNPNQYHYV A•RICRRFPF MKH•NVPSEQ EAHLCHFWDS D•SSCGHWKA
 651    KEGRWLLGHL SQGSRLQPAD VSFSLQFFL• FF•AFKLFLL LCAK•L•DWT
 701    RRCY•REVPL KGLVLVSKSC KFGLFSCVIM SAQ•RRP•ML HFLALKIP•V
 751    SLKTARSDDQ HLE•KQGPFM GEH•KEPGFK AGEWMTHPSH WPLPVSCISK
 801    SCKLWWLIFR KSGF•VSSLR CQGHGVRPAA SVRQLSSL•I SVLGVGASVF
 851    VFPF•D•VWQ SLFFCIGVTA L•FFLIAVFV •LQ••SLVWF LQSCAGTILV
 901    LCCKL•KVYG DLKS•CCEAE VILWKD•KDF VGTWFCVVYI YMRLNSERKV
 951    Q••C•KGNYD KDTFEITFKS TLYFT••HVS F•LKATKGIL IMA•VFKAIF
1001    SGIYQVYII• FCAKLLRVSF •NMRV•NMTP CGFPY•NPHS LIVIFIFENF
1051    HL•VP•YVV• ERPNRFLFFF LISSLCLEIV NIVI•CRLTL NKISLIGLKI
1101    TLIKLCDMQM TH
```

```
                                  MYQLFSR VRGGQQALLQ HWSEYIKTFG
 201    TAIVINPEKD KDMVQDLLDF KDKVDHVIEV CFQKNERFVN LMKESFETFI
 251    NKRPNKPAEL IAKHVDSKLR AGNKEATDEE LERTLDKIMI LFRFIHGKDV
 301    FEAFYKKDLA KRLLVGKSAS VDAEKSMLSK LKHECGAAFT SKLEGMFKDM
 351    ELSKDIMVHF KQHMQNQSDS GPIDLTVNIL TMGYWPTYTP MEVHLTPEMI
 401    KLQEVFKAFY LGKHSGRKLQ WQTTLGHAVL KAEFKEGKKE FQVSLFQTLV
 451    LLMFNEGDGF SFEEIKMATG IEDSELRRTL QSLACGKARV LIKSPKGKEV
 501    EDGDKFIFNG EFKHKLFRIK INQIQMKETV EEQVSTTERV FQDRQYQIDA
 551    AIVRIMKMRK TLGHNLLVSE LYNQLKFPVK PGDLKKRIES LIDRDYMERD
 601    KDNPNQYHYV A
```

Figure 18

+ strand (sense)	sequence (5'-->3')	
	1st base	
1. pch14-sp6-1f	686	GGC TTA ACA CTC AAT GTA C
2. pch14-sp6-2f	1005	CTA TGA AAA GAC AGC TTA AG
3. pch14-SP6-3f	1315	ATT TAG TTT GAA AAG CAT G
4 pch14-sp6-4f	1589	CAG ACT TTA AAG TCA CAA G
5. pch14-sp6-5f	1808	CAA AGA CTT GGT GTA TAG TG
- strand(antisense)	sequence (5'-->3')	
6. pch14-sp6-6fb	2020	GCA GTT TAA TTT GGT CCT G
7. pch14-sp6-5fb	1757	CTG TAA TTA TAG TTC TGT C
8. pch14-sp6-4fb	1607	CTT GTG ACT TTA AAG TCT G
9. pch14-sp6-3fb	1339	ATA ATC ATG CTT TTC AAA C
10.pch14-sp6-2rb	1023	TTA AGC TGT CTT TTC ATA G
11.pch14-sp6-1rb	704	GTA CAT TGA GTG TTA AAC C
12. CH14a	629	CGG CAG AGC TGA CTA CTG GAA GG
13. CH14b	644	CAA GCA GGG AAG TAA CGG CAG
14. CH14c	109	CTT GTT AGC TTG TTT AGA AGG TGG AAG AG
15. CH14d	90	GGT GGA AGA GAA GGT CTC CTT TCA GGC

Figure 19

```
   1  GAAGATGATG ATTACGGGTC TCGAACAGGA AGCATCTCCA GCAGTGTGTC
  51  TGTGCCTGCA AAGCCTGAAA GGAGACCTTC TCTTCCACCT TCTAAACAAG
 101  CTAACAAGAA TCTGATTTTG AAGGCTATAT CTGAAGCTCA AGAATCCGTA
 151  ACAAAAACAA CTAACTACTC TACAGTTCCA CAGAAACAGA CACTTCCAGT
 201  TGCTCCCAGA ACTCGAACTT CTCAAGAAGA ATTGCTAGCA GAAGTGGTCC
 251  AGGGACAAAG TAGGACCCCC AGAATAAGTC CCCCCATTAA AGAAGAGGAA
 301  ACAAAAGGAG ATTCTGTAGA AAAAAATCAA GCTGAGATGA GTGAACTGAG
 351  TGTGGCACAG AAACCAGAAA AACTTTTGGA GCGCTGCAAG TACTGGCCTG
 401  CTTGTAAAAA TGGGGATGAG TGTGCCTACC ATCACCCCAT CTCACCCTGC
 451  AAAGCCTTCC CCAATTGTAA ATTTGCTGAA AAATGTTTGT TTGTTCACCC
 501  AAATTGTAAA TATGATGCAA AGTGTACTAA ACCAGATTGT CCCTTCACTC
 551  ATGTGAGTAG AAGAATTCCA GTACTGTCTC CAAAACCAGT TGCACCACCA
 601  GCACCACCTT CCAGTAGTCA GCTCTGCCGT TACTTCCCTG CTTGTAAGAA
 651  GATGGAATGT CCCTTCTATC ATCCAAAACA TTGTAGGTTT AACACTCAAT
 701  GTACAAGTCC GGACTGCACA TTCTACCATC CCACCATTAA TGTCCCACCA
 751  CGACATGCCT TGAAATGGAT TCGACCTCAA ACCAGCGAAT AGCACCCAGT
 801  CCTGCCTGGC AGAAGATCAT GCAGTTTGGA AGTTTTCATG TACTGATGAA
 851  AGATACTCTA CAGAACTTGT CAAATCTTTG AAACTTGGAA TATATTGCTT
 901  TCATAATATG AAGTTTTATT GCCTATCTAT CTGAAGTGTC TAATTTTTCA
 951  AGTTTGTAAG TTTATTATGT GGTTTTAACA TTGGGTGTTT TTGTTTTGTT
1001  TTTACTATGA AAAGACAGCT TAAGGAAGAG CTAAATTCTG TTAAAATATT
1051  TGGGGCATGT TTGTGCACTG CTGTTGTGAG GATCAGCATA TGAAATTGAC
1101  ATCATGGTTA GTCATGGTAC TGCAGCTTAG GGGGCTACAC GGTTGCTGTG
1151  TGAGTGGAGA GATGCAGTGA GGCAGTTGTC ATTATTCTAA AAATTGTACT
1201  ACTTTCACTT TTCCCAAAGA TTATATAATG TTCATAATCC ACCATGAAAA
1251  CAGCATTGGC CAAAGGTACT GAGGCTGCTT AAAATATTCA ATTCTGCTTT
1301  TTAGTTTTTA AGTGAATTTA GTTTGAAAAG CATGATTATA CAGGCCTCTC
1351  GAGGCTGAGT GCTACTTTCG GTAAAGTTCC AGTTTTCCAG CCTTCTGTGA
1401  CAGGATGAAT GAGGTGGGTA TGGACAGTGG AGGCAGCTGG AATGGCAAGT
1451  GCAGAAAATA GGAACAGTTC TATACAGTGC TCTCATTTAC TAATAACATA
1501  ATGCCTTCTA AATAATTTTT TTGGGAAACT ACATTATCAC AAAATTATAC
1551  AAATTTTTTT ACAAGTATTT ACATACTGTA TCTGAAAACA GACTTTAAAG
1601  TCACAAGATT ATAAATGTAC ATATGTATTC TCACATTCTG AAAAATAACA
1651  TTCTCAGAAT CCACAGAAAA TATACTTAGT TACTACTGAA GATAATTTTT
1701  GAAATGTAAA AATTAGATTT AAATAGTATA TTTTAAATGA CAGAACTATA
1751  ATTACAGAGA TCAGATCAGA TAGGTAAACT GCAAGATAGA TAGGATGAAA
1801  CTTTTGGCCT ACTGTATTAC TTACAGAGTT TTTTTGTGTG TGGTTTTTAA
1851  AACTGTTAAG GCAAGAAGTG TCAAATGCTT TAGAGTTAAA TAACAGATCA
1901  CTGATTTCAA AGACTTGGTG TATAGTGTTA AAAATTAAAG CTTAAAAGGT
1951  GGTTAGAAAA GTGGATTAAT GCAAAAGGGG TAATAAAGAC TGCAACATTC
2001  TCAGGACCAA ATTAAACTGC T
```

Figure 20

```
  1  EDDDYGSRTG SISSSVSVPA KPERRPSLPP SKQANKNLIL KAISEAQESV
 51  TKTTNYSTVP QKQTLPVAPR TRTSQEELLA EVVQGQSRTP RISPPIKEEE
101  TKGDSVEKNQ AEMSELSVAQ KPEKLLERCK YWPACKNGDE CAYHHPISPC
151  KAFPNCKFAE KCLFVHPNCK YDAKCTKPDC PFTHVSRRIP VLSPKPVAPP
201  APPSSSQLCR YFPACKKMEC PFYHPKHCRF NTQCTSPDCT FYHPTINVPP
251  RHALKWIRPQ TSE•HPVLPG RRSCSLEVFM Y••KILYRTC QIFETWNILL
301  S•YEVLLPIY LKCLIFQVCK FIMWF•HWVF LFCFYYEKTA •GRAKFC•NI
351  WGMFVHCCCE DQHMKLTSWL VMVLQLRGLH GCCVSGEMQ• GSCHYSKNCT
401  TFTFPKDYIM FIIHHENSIG QRY•GCLKYS ILLFSF•VNL V•KA•LYRPL
451  EAECYFR•SS SFPAFCDRMN EVGMDSGGSW NGKCRK•EQF YTVLSFTNNI
501  MPSK•FFWET TLSQNYTNFF TSIYILYLKT DFKVTRL•MY ICILTF•KIT
551  FSESTENILS YY•R•FLKCK N•I•IVYFK• QNYNYRDQIR •VNCKIDRMK
601  LLAYCITYRV FLCVVFKTVK ARSVKCFRVK •QITDFKDLV YSVKN•SLKG
651  G•KSGLMQKG ••RLQHSQDQ IKL
```

```
EDDDYGSRTG SISSSVSVPA KPERRPSLPP SKQANKNLIL KAISEAQESV
TKTTNYSTVP QKQTLPVAPR TRTSQEELLA EVVQGQSRTP RISPPIKEEE
TKGDSVEKNQ AEMSELSVAQ KPEKLLERCK YWPACKNGDE CAYHHPISPC
KAFPNCKFAE KCLFVHPNCK YDAKCTKPDC PFTHVSRRIP VLSPKPVAPP
APPSSSQLCR YFPACKKMEC PFYHPKHCRF NTQCTSPDCT FYHPTINVPP
RHALKWIRPQ TSE
```

Figure 21

```
  1    AAAACTTTCG GAAGAGAAAG TTGCCTGTGG TAAGTTCAGT TGTTAAAGTA
 51    AAAAAATTCA ATCATGATGG AGAAGAGGAG GAAGAAGATG ATGATTACGG
101    GTCTCGAACA GGAAGCATCT CCAGCAGTGT GTCTGTGCCT GCAAAGCCTG
151    AAAGGAGACC TTCTCTTCCA CCTTCTAAAC AAGCTAACAA GAATCTGATT
201    TTGAAGGCTA TATCTGAAGC TCAAGAATCC GTAACAAAAA CAACTAACTA
251    CTCTACAGTT CCACAGAAAC AGACACTTCC AGTTGCTCCC AGAACTCGAA
301    CTTCTCAAGA AGAATTGCTA GCAGAAGTGG TCCAGGGGAC AAAGTAGGAC
351    CCCCAGAATA AGTCCCCCCA TTAAAGAAGA GGAAACAAAA GGAGATTCTG
401    TAGAAAAAAA TCAAGATTAC TATGACATGG AATCCATGGT CCATGCAGAC
451    ACAAGATCAT TTATTCTGAA GAAGCCAAAG CTGTCTGAGG AAGTANTAGT
501    GGCACCAAAC CAAGANTCGG GGATGAAGAC TGCAGATTCC CTTCGGGTTC
551    TTTCAGGGAC CCTTATGCAG ACACNAGATC TTGTTCAACC AGATAAACCT
601    GCAAGTCCCA AG
```

```
  1    KTFGRESCLW •VQLLK•KNS IMMEKRRKKM MITGLEQEAS PAVCLCLQSL
 51    KGDLLFHLLN KLTRI•F•RL YLKLKNP•QK QLTTLQFHRN RHFQLLPELE
101    LLKKNC•QKW SRGQSRTPRI SPPIKEEETK GDSVEKNQDY YDMESMVHAD
151    TRSFILKKPK LSEEVXVAPN QXSGMKTADS LRVLSGTLMQ TXDLVQPDKP
201    ASPK
```

```
  1    NAGCTGCTCT GACGGGNAGN GGAATGNATG GNGGCTTGTT CNGAAACNNG
 51    CCAGATGGCG NGAGGGGGAC AAGTAGCGGC GTGATTNAGA AGAGGGAGGT
101    GAGGGTNCTC ACATCACCNC ATCTNACCAT GNCGNGCCNT CCCCANTANT
151    AANANTGATG ATAGNGGGAA GTGGGCCCAC CCAGAAGCNT GATTGAGCGG
201    CCGCCAGTAN GAAACNNGTT TGTCCANTTA GNCATACNNA TNGTAGGGTT
251    CNAGCNGCGT CCCCGGCACC NGCANANNNN CNNCNGGGAC NACNGCCCNN
301    NNNTNNGTTA NNCNGNGNAG NNAAAAAATT CAATCATGAT GGAGAAGAGG
351    AGGAAGAAGA TGATGATTAC GGGTCTCGAA CAGGAAGCAT CTCCAGCAGT
401    GTGTCTGTGC CTGCAAA
```

Untitled translated in RF 2

```
  1    SCSDGXXNXW XLVXKXARWX EGDK•RRDXE EGGEGXHITX SXHXXXSPXX
 51    XXMIXGSGPT QKXD•AAASX KXVCPXXHXX XRVXXASPAX AXXXXGXXPX
101    XXLXXXXKKF NHDGEEEEED DDYGSRTGSI SSSVSVPA
```

Figure 22

CH1-9a11-2

GA AAA CAA ATG GAA GAA ATG CAA AAG GCT TTC AAT AAA ACA ATC GTG
AAA CTT CAG AAT ACT TCA AGA ATA GCA GAG GAG CAG GAT CAG CGG CAA
ACT GAA GCC ATC CAG TTG CTA CAG GCA CAG CTG ACC AAC ATG ACA CAG
CTT GTT CAA

Lys Gln Met Glu Glu Met Gln Lys Ala Phe Asn Lys Thr Ile Val Lys
Leu Gln Asn Thr Ser Arg Ile Ala Glu Glu Gln Asp Gln Arg Gln Thr
Glu Ala Ile Gln Leu Leu Gln Ala Gln Leu Thr Asn Met Thr Gln Leu
Val Gln

CH8-2a13-1

GAA CAG GCA AGC AGA TAT GCT ACT GTC AGT GAA AGA GTG CAT GCT CAA
GTG CAG CAA TTT CTA AAA GAA GGT TAT TTA AGG GAG GAG ATG GTT CTG
GAC AAT ATC CCA AAG CTT CTG AAC TGC CTG AGA GAC TGC AAT GTT GCC
ATC CGA TGG CTG ATG CTT C

Glu Gln Ala Ser Arg Tyr Ala Thr Val Ser Glu Arg Val His Ala Gln
Val Gln Gln Phe Leu Lys Glu Gly Tyr Leu Arg Glu Glu Met Val Leu
Asp Asn Ile Pro Lys Leu Leu Asn Cys Leu Arg Asp Cys Asn Val Ala
Ile Arg Trp Leu Met Leu

CH13-2a12-1

CTC ACA ATG GGC TAC TGG CCA ACA TAC ACG CCC ATG GAA GTG CAC TTA
ACC CCA GAA ATG ATT AAA CTT CAG GAA GTA TTT AAG GCA TTT TAT CTT
GGA AAG CAC AG

Leu Thr Met Gly Tyr Trp Pro Thr Tyr Thr Pro Met Glu Val His Leu
Thr Pro Glu Met Ile Lys Leu Gln Glu Val Phe Lys Ala Phe Tyr Leu
Gly Lys His

CH14-2a16-1

TG TTT GTT CAC CCA AAT TGT AAA TAT GAT GCA AAG TGT ACT AAA CCA
GAT TGT CCC TTC ACT CAT GTG AGT AGA AGA ATT CCA GTA CTG TCT CCA
AAA CCA GTT GCA CCA CCA G

Phe Val His Pro Asn Cys Lys Tyr Asp Ala Lys Cys Thr Lys Pro Asp
Cys Pro Phe Thr His Val Ser Arg Arg Ile Pro Val Leu Ser Pro Lys
Pro Val Ala Pro Pro

Figure 23(A)

CTCAGAGAGG GCTGCCAGGA CGCGAGCCAC TGAGGAGCCG CTCAGCCAGC
GCCATAGCCC TTAGGACTAT CGGTCACATT CTCGCGCTCC TGCTCCGGCT
CCTCCATCTT GGCCTCGGCA GTGGCGGCTG CCGGGAGGAT GTGCCGCCTT
CTGGCAGGGG GAAGAAGGAG GAGAAGATGA AGAAGCACCG GCGGGCCTTG
GCCCTGGTCT CCTGCCTCTT TCTGTGCTCT CTGGTCTGGC TTCCCAGCTG
GCGTGTATGT TGTAAAGAGA GTTCCTCAGC TTCAGCGTCA TCATATTACT
CTCAAGATGA CAACTGCGCA CTAGAAAATG AAGATGTACA ATTCCAGAAA
AAGAATACAG AGTCAAAAAA GTTAAGTCCA CCGGTGGTGG AGACACTCCC
TACAGTTGAT TTGCATGAAG AGTCTTCCAA TGCAGTTGTG GACAGTGAAA
CTGTTGAAAA TATTTCCAGC TCATCTACCT CAGAAATCAC TCCAATCTCA
AAGCTTGATG AAATAGAAAA ATCTGGTACT ATTCCGATAG CCAAACCAAG
TGAAACTGAG CAGTCTGAAA CTGATTGTGA TGTTGGTGAG GCCCTTGATG
CTAGTGCTCC AATTGAACAA CCTTCCTTTG TCAGTCCACC TGACAGCCTT
GTTGGCCAGC ATATAGAAAA TGTATCATCT TCACATGGTA AAGGAAAGAT
AACAAAATCA GAATTTGAAT CAAAAGTTTC AGCAAGTGAA CAGGGCGGTG
GTGATCCAAA ATCTGCATTG AATGCTTCAG ATAATTTAAA AAATGAGAGC
TCTGATTATA CAAAACCAGG AGACATTGAC CCTACATCAG TAGCAAGTCC
CAAAGATCCA GAAGATATAC CAACATTTGA TGAATGGAAG AAGAAAGTTA
TGGAAGTAGA AAAAGAAAAA AGTCAGTCGA TGCATGCATC TTCTAATGGA
GGTTCACATG CCACCAAAAA GGTCCAGAAA AATCGAAATA ATTATGCCTC
AGTAGAATGT GGTGCCAAAA TTCTAGCAGC TAATCCAGAA GCCAAGAGCA
CATCTGCTAT TCTTATAGAA AATATGGATC TTTACATGTT GAATCCTTGC
AGCACTAAAA TTTGGTTTGT TATTGAACTT TGTGAACCAA TTCAAGTAAA
ACAGCTTGAT ATTGCAAATT ATGAATTATT TTCTTCTACT CCTAAAGATT
TTCTGGTTTC TATCAGTGAC AGATATCCAA CAAATAAGTG GATTAAGCTG
GGTACTTTTC ATGGTAGAGA TGAGCGGAAT GTACAGAGTT TCCCTTTAGA
TGAACAGATG TATGCAAAAT ATGTCAAGGT TGAGTTGCTA TCACATTTTG
GATCAGAGCA CTTTTGTCCA TTAAGCCTTA TAAGGGTATT TGGCACTAAC
ATGGTGGAAG AATATGAAGA AATTGCTGAT TCCCAGTATC ACTCAGAACG
CCAGGAACTA TTTGATGAGG ACTATGATTA TCCACTGGAT TATAATACTG
GAGAGGATAA ATCCTCAAAA AATCTTCTTG GTTCTGCTAC AAATGCCATT
CTAAATATGG TGAATATTGC TGCTAATATT CTGGGAGCAA AAACTGAAGA
CCTGACAGAA GGAAATAAAA GTATATCTGA GAATGCCACT GCCACAGCTG
CACCTAAAAT GCCTGAATCA ACTCCTGTTT CAACTCCTGT TCCATCTCCT
GAGTATGTAA CCACTGAAGT ACACACACAT GACATGGAGC CGTCAACACC
AGATACTCCA AAAGAGAGTC CCATTGTACA GTTAGTTCAA GAGGAGGAAG
AGGAGGCAAG TCCATCTACA GTGACCCTTC TGGGCAGCGG TGAACAGGAA
GATGAATCAT CACCCTGGTT TGAGTCAGAG ACACAAATAT TTTGCAGTGA

Figure 23(B)

ACTGACCACA ATTTGTTGTA TTTCTAGTTT TTCAGAATAC ATATATAAAT
GGTGTTCAGT TAGAGTTGCT CTTTATCGGC AGCGCAGCCG AACTGCTTTG
AGTAAAGGAA AAGATTATCT TGTGTTAGCT CAACCACCCT TACTACTTCC
TGCGGAATCA GTAGATGTTT CAGTATTGCA ACCTCTGAGT GGAGAATTGG
AAAATACGAA TATAGAAAGG GAAGCTGAAA CTGTTGTTCT GGGTGATTTA
AGTAGTAGTA TGCACCAGGA TGACTTGGTG AATCACACTG TAGATGCAGT
TGAACTTGAA CCAAGCCATT CTCAAACTCT TTCTCAGTCT CTTCTTTTAG
ATATTACCCC AGAAATCAAT CCCTTGCCTA AAATAGAAGT ATCTGAGTCT
GTTGAATATG AGGCAGGACA TATACCATCA CCAGTGATTC CCCAAGAGAG
TTCTGTTGAG ATCGATAATG AAACAGAACA AAAGTCTGAG AGCTTTAGTT
CTATAGAGAA ACCATCTATT ACCTATGAAA CAAATAAAGT TAATGAGTTA
ATGGATAATA TTATAAAAGA AGATATGAAC TCCATGCAAA TTTTCACAAA
GCTGTCTGAA ACAATAGTGC CACCAATAAA TACAGCCACT GTACCCGACA
ATGAAGATGG GGAAGCCAAA ATGAATATAG CTGACACAGC AAAGCAAACT
TTGATTTCTG TTGTGGATTC TTCTTCATTA CCTGAAGTAA AAGAAGAAGA
ACAGTCTCCA GAAGATGCCC TTTTGAGAGG GTTACAGAGG ACAGCTACAG
ATTTTTATGC TGAATTGCAA AATTCTACAG ATCTAGGATA TGCTAATGGA
AATCTTGTAC ATGGATCAAA CCAAAAGGAG TCAGTATTTA TGAGACTTAA
TAATCGTATT AAAGCCTTAG AAGTTAACAT GTCTCTCAGT GGTCGCTATC
TGGAGGAGCT TAGCCAAAGG TACCGAAAAC AAATGGAAGA AATGCAAAAG
GCTTTCAACA AAACAATCGT GAAACTTCAG AATACTTCAA GAATAGCAGA
GGAGCAGGAT CAGCGGCAAA CTGAAGCCAT CCAGTTGCTA CAGGCACAGC
TGACCAACAT GACACAGCTT GTTTCAAATT TATCAGCAAC AGTAGCAGAA
TTGAAACGGG AGGTTTCAGA TCGACAAAGC TATCTTGTCA TATCTTTGGT
TCTTTGTGTT GTCTTGGGAC TGATGCTTTG TATGCAGCGT TGTCGAAATA
CTTCTCAATT TGATGGAGAT TATATTTCAA AACTTCCTAA AAGTAATCAG
TATCCAAGCC CTAAAAGGTG TTTCTCTTCC TATGATGATA TGAATTTGAA
AAGAAGAACT TCATTCCCAC TCATGAGATC CAAGTCTCTA CAGTTAACTG
GCAAAGAAGT AGACCCAAAT GATTTGTACA TTGTAGAACC CCTCAAGTTT
TCTCCAGAAA AGAAGAAGAA GCGCTGCAAG TACAAAATTG AAAAAATTGA
GACCATAAAG CCTGAAGAAC CATTGCACCC CATAGCCAAT GGCGACATAA
AAGGAAGAAA GCCCTTTACG AACCAGAGAG ATTTTTCTAA TATGGGAGAA
GTTTATCACT CTTCTTATAA AGGTCCTCCA TCTGAAGGAA GCTCAGAAAC
TTCATCACAG TCAGAAGAGT CCTATTTTTG TGGCATTTCA GCTTGCACAA
GTCTGTGCAA TGGACAGTCT CAAAAGACAA AAACTGAGAA GAGGGCTTTA
AAACGAAGAC GATCTAAAGT CCAAGACCAA GGAAAATTGA TAAAAACTCT
AATACAGACT AAGTCGGGAT CATTGCCGAG CCTGCATGAC ATAATCAAAG
GAAACAAAGA GATCACCGTG GGAACATTTG GTGTTACAGC AGTCTCGGGA

Figure 23(C)

CATATCTAAA ATTAATTGAA CTTTTCATAC AGAAGACTTT TTTGTTGTTG
TTCTTTGAAG AACAGTCTGT AGTATTTGAA GGGTTTGGGG GAGGGAGAAA
ATATTAATGG GAAAGGCATT CAGAAATTAT GGTTTCTACC TTTTTAAAAA
GTAGATGGGA TTGTGCTCAA TCTTGGTTAA TGAGCTACAG TTTTACAAAG
CTGATCACTT CCTATAAGGA CAATGGTAGA CATTTTATAA AGATGTTTTT
TCACAAGATT AATTACTGGG ACAAAAGTAA TTTGGAAGCC CAGTTCCTTA
GGTGGGATAG GAATGAAAGC CTAAACCTCT TCCTTTAGCT TTGTTCCTAT
TTCTTGCACC TTCCCATATT TATGTGCCTT TTGTCTATTT ATAATGCCAC
TGGAAGAGGA GGGATAACTT TTTCTGTTAT TTGATTTCTT TTATAACTTT
GTTAGGTTTT TGAAGCTGCA AACACTACAA TGCTTTGAGG GGGTCTGTGC
CTGAAGCTCA GGAGTGTGGA TCAGACAGTC TAAAGATCCT AAAAACTTGC
CAACTGGATC TTTGTTTAGC AAACTCACTG GAAATGAACA CTTAATGGAA
TTTTTAAGTC TGTTCTGTTA GGTAGATGGT GATGCTCTTG TTATTTTCAC
TTATTCAGGC TGGATTACTT CTTACTTAGT TACTAACTCA ATGAGGAAAA
AATCCCTACA GGATCTTTTT TTGCAAACAA CTGATATATG CAGACAAATT
TTTGACAAAT TCACCTTTTA AACACGACGT TAACCGATTT GTGAAGGTTT
TCTTTAGCTT ACATTTTAAA CATACACAAT AAACACTAAT CCTCCAAACT
TTCACTGTTT TTATTAGTAT GAATATAAAA TTTGAAGGTT TGGCCAATTA
GTACAAGTCT CATGATATAA TCACAGCCTG CATACATATG CACAGATCCA
GTTAGTGAGT TTGTCAAGCT TAATCTAATT GGTTAAGTCT AAAGAGATTA
TTATTCCTTG ATGTTTGCTT TGTATTGGCT ACAAATGTGC AGAGGTAATA
CATATGTGAT GTCGATGTCT CTGTCTTTTT TTTTGTCTTT AAAAAATAAT
TGGCAGCAAC TGTATTTGAA TAAAATGATT TCTTAGTATG ATTGTACAGT
AATGAATGAA AGTGGAACAT GTTTCTTTTT GAAAGGGAGA GAATTGACCA
TTTATTGTTG TGATGTTTAA GTTATAACTT ATTGAGCACT TTTAGTAGTG
ATAACTGTTT TTAAACTTGC CTAATACCTT TCTTGGGTAT TGTTTGTAAT
GTGACTTATT TAACGCCTTC TTTGTTTGTT TAAGTTGCTG CTTTAGGTTA
ACAGCGTGTT TTAGAAGATT TAAATTTCTT TCCTGTCTGC ACAATTAGCT
ATTCAGAGCA AGAGGGCCTG ATTTTATAGA AGCCCCTTGA AAAGAGGTCC
AGATGAGAGC AGAGATACAG TGAGAAATTA TGTGATCTGT GTGTTGTGGG
AAGAGAATTT TCAATATGTA ACTACGGAGC TGTAGTGCCA TTAGAAACTG
TGAATTTCCA AATAAATCTG AACACTTGTC TTTATT

Figure 24

QRGLPGREPL RSRSASAIAL RTIGHILALL LRLLHLGLGS GGCREDVPPS
GRGKKEEKMK KHRRALALVS CLFLCSLVWL PSWRVCCKES SSASASSYYS
QDDNCALENE DVQFQKKNTE SKKLSPPVVE TLPTVDLHEE SSNAVVDSET
VENISSSSTS EITPISKLDE IEKSGTIPIA KPSETEQSET DCDVGEALDA
SAPIEQPSFV SPPDSLVGQH IENVSSSHGK GKITKSEFES KVSASEQGGG
DPKSALNASD NLKNESSDYT KPGDIDPTSV ASPKDPEDIP TFDEWKKKVM
EVEKEKSQSM HASSNGGSHA TKKVQKNRNN YASVECGAKI LAANPEAKST
SAILIENMDL YMLNPCSTKI WFVIELCEPI QVKQLDIANY ELFSSTPKDF
LVSISDRYPT NKWIKLGTFH GRDERNVQSF PLDEQMYAKY VKVELLSHFG
SEHFCPLSLI RVFGTNMVEE YEEIADSQYH SERQELFDED YDYPLDYNTG
EDKSSKNLLG SATNAILNMV NIAANILGAK TEDLTEGNKS ISENATATAA
PKMPESTPVS TPVPSPEYVT TEVHTHDMEP STPDTPKESP IVQLVQEEEE
EASPSTVTLL GSGEQEDESS PWFESETQIF CSELTTICCI SSFSEYIYKW
CSVRVALYRQ RSRTALSKGK DYLVLAQPPL LLPAESVDVS VLQPLSGELE
NTNIEREAET VVLGDLSSSM HQDDLVNHTV DAVELEPSHS QTLSQSLLLD
ITPEINPLPK IEVSESVEYE AGHIPSPVIP QESSVEIDNE TEQKSESFSS
IEKPSITYET NKVNELMDNI IKEDMNSMQI FTKLSETIVP PINTATVPDN
EDGEAKMNIA DTAKQTLISV VDSSSLPEVK EEEQSPEDAL LRGLQRTATD
FYAELQNSTD LGYANGNLVH GSNQKESVFM RLNNRIKALE VNMSLSGRYL
EELSQRYRKQ MEEMQKAFNK TIVKLQNTSR IAEEQDQRQT EAIQLLQAQL
TNMTQLVSNL SATVAELKRE VSDRQSYLVI SLVLCVVLGL MLCMQRCRNT
SQFDGDYISK LPKSNQYPSP KRCFSSYDDM NLKRRTSFPL MRSKSLQLTG
KEVDPNDLYI VEPLKFSPEK KKKRCKYKIE KIETIKPEEP LHPIANGDIK
GRKPFTNQRD FSNMGEVYHS SYKGPPSEGS SETSSQSEES YFCGISACTS
LCNGQSQKTK TEKRALKRRR SKVQDQGKLI KTLIQTKSGS LPSLHDIIKG
NKEITVGTFG VTAVSGHI•N •LNFSYRRLF CCCSLKNSL• YLKGLGEGEN
INGKGIQKLW FLPF•KVDGI VLNLG••ATV LQS•SLPIRT MVDIL•RCFF
TRLITGTKVI WKPSSLGGIG MKA•TSSFSF VPISCTFPYL CAFCLFIMPL
EEEG•LFLLF DFFYNFVRFL KLQTLQCFEG VCA•SSGVWI RQSKDPKNLP
TGSLFSKLTG NEHLMEFLSL FC•VDGDALV IFTYSGWITS YLVTNSMRKK
SLQDLFLQTT DICRQIFDKF TF•TRR•PIC EGFL•LTF•T YTINTNPPNF
HCFY•YEYKI •RFGQLVQVS •YNHSLHTYA QIQLVSLSSL I•LVKSKEII
IP•CLLCIGY KCAEVIHM•C RCLCLFFCL• KIIGSNCI•I K•FLSMIVQ•
•MKVEHVSF• KGEN•PFIVV MFKL•LIEHF ••••LFLNLP NTFLGYCL•C
DLFNAFFVCL SCCFRLTACF RRFKFLSCLH N•LFRARGPD FIEAP•KEVQ
MRAEIQ•EIM •SVCCGKRIF NM•LRSCSAI RNCEFPNKSE HLSL

Figure 25(A)

TAGAATTCAG CGGCCGCTGA ATTCTAGCTG CGGGGTAGGA GTCCGCGGCA
GCCTCCGGGT AAGCCAAGCG CCGCGCAGTG CTGAGTTCCC GCACGCCGCA
GAGCCATGGA GATCGGCACC GAGACCAGCC GCAAGATCCG GAGTGCCATT
AAGGGGAAAT TACAAGAATT AGGAGCTTAT GTTGATGAAG AACTTCCTGA
TTACATTATG GTGATGGTGG CCAACAAGAA AAGTCAGGAC CAAATGACAG
AGGATCTGTC CCTGTTTCTA GGGAACAACA CAATTCGATT CACCGTATGG
CTTCATGGTG TATTAGATAA ACTTCGCTCT GTTACAACTG AACCCTCTAG
TCTGAAGTCT TCTGATACCA ACATCTTTGA TAGTAACGTG CCTTCAAACA
AGAACAATTT CAGTCGGGGA GATGAGAGGA GGCATGAAGC TGCAGTGCCA
CCACTTGCCA TTCCTAGCGC GAGACCTGAA AAAAGAGATT CCAGAGTTTC
TACAAGTTCG CAGGAGTCAA AAACCACAAA TGTCAGACAG ACTTACGATG
ATGGAGCTGC AACCCGACTA ATGTCAACAG TGAAACCTTT GAGGGAGCCA
GCACCCTCTG AAGATGTGAT TGATATTAAG CCAGAACCAG ATGATCTCAT
TGACGAAGAC CTCAACTTTG TGCAGGAGAA TCCCTTATCT CAGAAAGAAC
CTACAGTGAC ACTTACATAT GGTTCTTCTC GCCCTTCTAT TGAAATTTAT
CGACCACCTG CAAGTAGAAA TGCAGATAGT GGTGTTCATT TAAACAGGTT
GCAATTTCAA CAGCAGCAGA ATAGTATTCA TGCTGCCAAG CAGCTTGATA
TGCAGAGTAG TTGGGTATAT GAAACAGGAC GTTTGTGTGA ACCAGAGGTG
CTTAACAGCT TAGAAGAAAC GTATAGTCCG TTCTTTAGAA ACAACTCGGA
GAAAATGAGT ATGGAGGATG AAAACTTTCG GAAGAGAAAG TTGCCTGTGG
TAAGTTCAGT TGTTAAAGTA AAAAAATTCA ATCATGATGG AGAAGAGGAG
GAAGGAGATG ATGATTACGG GTCTCGAACA GGAAGCATCT CCAGCAGTGT
GTCTGTGCCT GCAAAGCCTG AAAGGAGACC TTCTCTTCCA CCTTCTAAAC
AAGCTAACAA GAATCTGATT TTGAAGGCTA TATCTGAAGC TCAAGAATCC
GTAACAAAAA CAACTAACTA CTCTACAGTT CCACAGAAAC AGACACTTCC
AGTTGCTCCC AGAACTCGAA CTTCTCAAGA AGAATTGCTA GCAGAAGTGG
TCCAGGGACA AAGTAGGACC CCCAGAATAA GTCCCCCCAT TAAAGAAGAG
GAAACAAAAG GAGATTCTGT AGAAAAAAAT CAAGCTGAGA TGAGTGAACT
GAGTGTGGCA CAGAAACCAG AAAAACTTTT GGAGCGCTGC AAGTACTGGC
CTGCTTGTAA AAATGGGGAT GAGTGTGCCT ACCATCACCC CATCTCACCC
TGCAAAGCCT TCCCCAATTG TAAATTTGCT GAAAAATGTT TGTTTGTTCA
CCCAAATTGT AAATATGATG CAAAGTGTAC TAAACCAGAT TGTCCCTTCA
CTCATGTGAG TAGAAGAATT CCAGTACTGT CTCCAAAACC AGTTGCACCA
CCAGCACCAC CTTCCAGTAG TCAGCTCTGC CGTTACTTCC CTGCTTGTAA
GAAGATGGAA TGTCCCTTCT ATCATCCAAA ACATTGTAGG TTTAACACTC
AATGTACAAG TCCGGACTGC ACATTCTACC ATCCCACCAT TAATGTCCCA
CCACGACATG CCTTGAAATG GATTCGACCT CAAACCAGCG AATAGCACCC
AGTCCTGCCT GGCAGAAGAT CATGCAGTTT GGAAGTTTTC ATGTACTGAT

Figure 25(B)

GAAAGATACT CTACAGAACT TGTCAAATCT TTGAAACTTG GAATATATTG
CTTTCATAAT ATGAAGTTTT ATTGCCTATC TATCTGAAGT GTCTAATTTT
TCAAGTTTGT AAGTTTATTA TGTGGTTTTA ACATTGGGTG TTTTTGTTTT
GTTTTTACTA TGAAAAGACA GCTTAAGGAA GAGCTAAATT CTGTTAAAAT
ATTTGGGGCA TGTTTGTGCA CTGCTGTTGT GAGGATCAGC ATATGAAATT
GACATCATGG TTAGTCATGG TACTGCAGCT TAGGGGGCTA CACGGTTGCT
GTGTGAGTGG AGAGATGCAG TGAGGCAGTT GTCATTATTC TAAAAATTGT
ACTACTTTCA CTTTTCCCAA AGATTATATA ATGTTCATAA TCCACCATGA
AAACAGCATT GGCCAAAGGT ACTGAGGCTG CTTAAAATAT TCAATTCTGC
TTTTTAATTT TTAAGTGAAT TTAGTTTGAA AAGCATGATT ATACAGGCCT
CTCAGGCTGA GTGCTACTTT CGGTAAAGTT CCAGTTTTCC TGCCTTCTGT
GACAGGATGA ATGAGGTGGG TATGGACAGT GGAGGCAGCT GGAATGGCAA
GTGCAGAAAA TAGGAACAGT TCTATACAGT GCTCTCATTT ACTAATAACA
TAATGCCTTC TAAATAATTT TTTTGGGAAA CTACATTATC ACAAAATTAT
ACAAATTTTT TTACAAGTAT TTACATACTG TATCTGAAAA CAGACTTTAA
AGTCACAAGA TTATAAATGT ACATATGTAT TCTCACATTC TGAAAAATAA
CATTCTCAGA ATCCACAGAA AATATACTTA GTTACTACTG AAGATAATTT
TTGAAATGTA AAAATTAGAT TTAAATAGTA TATTTTAAAT GACAGAACTA
TAATTACAGA GATCAGATCA GATAGGTAAA CTGCAAGATA GATAGGATGA
AACTTTTGGC CTACTGTATT ACTTACAGAG TTTTTTTGTG TGTGGTTTTT
AAAACTGTTA AGGCAAGAAG TGTCAAATGC TTTAGAGTTA AATAACAGAT
CACTGATTTC AAAGACTTGG TGTATAGTGT TAAAAATTAA AGCTTAAAAG
GTGGTTAGAA AAGTGGATTA ATGCAAAAGG GGTAATAAAG ACTGCAACAT
TCTCAGGACC AAATTAAACT GCTAA

Figure 26

```
•NSAAAEF•L RGRSPRQPPG KPSAAQC•VP ARRRAMEIGT ETSRKIRSAI
KGKLQELGAY VDEELPDYIM VMVANKKSQD QMTEDLSLFL GNNTIRFTVW
LHGVLDKLRS VTTEPSSLKS SDTNIFDSNV PSNKNNFSRG DERRHEAAVP
PLAIPSARPE KRDSRVSTSS QESKTTNVRQ TYDDGAATRL MSTVKPLREP
APSEDVIDIK PEPDDLIDED LNFVQENPLS QKEPTVTLTY GSSRPSIEIY
RPPASRNADS GVHLNRLQFQ QQQNSIHAAK QLDMQSSWVY ETGRLCEPEV
LNSLEETYSP FFRNNSEKMS MEDENFRKRK LPVVSSVVKV KKFNHDGEEE
EGDDDYGSRT GSISSSVSVP AKPERRPSLP PSKQANKNLI LKAISEAQES
VTKTTNYSTV PQKQTLPVAP RTRTSQEELL AEVVQGQSRT PRISPPIKEE
ETKGDSVEKN QAEMSELSVA QKPEKLLERC KYWPACKNGD ECAYHHPISP
CKAFPNCKFA EKCLFVHPNC KYDAKCTKPD CPFTHVSRRI PVLSPKPVAP
PAPPSSSQLC RYFPACKKME CPFYHPKHCR FNTQCTSPDC TFYHPTINVP
PRHALKWIRP QTSE•HPVLP GRRSCSLEVF MY••KILYRT CQIFETWNIL
LS•YEVLLPI YLKCLIFQVC KFIMWF•HWV FLFCFYYEKT A•GRAKFC•N
IWGMFVHCCC EDQHMKLTSW LVMVLQLRGL HGCCVSGEMQ •GSCHYSKNC
TTFTFPKDYI MFIIHHENSI GQRY•GCLKY SILLFNF•VN LV•KA•LYRP
LRLSATFGKV PVFLPSVTG• MRWVWTVEAA GMASAENRNS SIQCSHLLIT
•CLLNNFFGK LHYHKIIQIF LQVFTYCI•K QTLKSQDYKC TYVFSHSEK•
HSQNPQKIYL VTTEDNF•NV KIRFK•YILN DRTIITEIRS DR•TAR•IG•
NFWPTVLLTE FFCVWFLKLL RQEVSNALEL NNRSLISKTW CIVLKIKA•K
VVRKVD•CKR GNKDCNILRT KLNC•
```

```
                                        MEIGT ETSRKIRSAI
KGKLQELGAY VDEELPDYIM VMVANKKSQD QMTEDLSLFL GNNTIRFTVW
LHGVLDKLRS VTTEPSSLKS SDTNIFDSNV PSNKNNFSRG DERRHEAAVP
PLAIPSARPE KRDSRVSTSS QESKTTNVRQ TYDDGAATRL MSTVKPLREP
APSEDVIDIK PEPDDLIDED LNFVQENPLS QKEPTVTLTY GSSRPSIEIY
RPPASRNADS GVHLNRLQFQ QQQNSIHAAK QLDMQSSWVY ETGRLCEPEV
LNSLEETYSP FFRNNSEKMS MEDENFRKRK LPVVSSVVKV KKFNHDGEEE
EGDDDYGSRT GSISSSVSVP AKPERRPSLP PSKQANKNLI LKAISEAQES
VTKTTNYSTV PQKQTLPVAP RTRTSQEELL AEVVQGQSRT PRISPPIKEE
ETKGDSVEKN QAEMSELSVA QKPEKLLERC KYWPACKNGD ECAYHHPISP
CKAFPNCKFA EKCLFVHPNC KYDAKCTKPD CPFTHVSRRI PVLSPKPVAP
PAPPSSSQLC RYFPACKKME CPFYHPKHCR FNTQCTSPDC TFYHPTINVP
PRHALKWIRP QTSE
```